Colicin translocation: investigating the topology of the translocon and potential applications

Yana Dem'yanenko

Linacre College



A thesis submitted for the degree of

Masters of Science by Research in Biochemistry

Supervisor: Professor Colin Kleanthous

Kleanthous Group

Department of Biochemistry, University of Oxford

South Parks Road, Oxfordshire, OX1 3QU

October 2016

Abstract

Colicins are proteins produced by some strains of *E. coli* to achieve competitive advantage over related species. With MW in the range between 30-75 kDa they consist of receptor, translocator and cytotoxic domains. After being released from producing cells, colicins cross the outer and, in some cases, inner membrane to kill susceptible cells. This ability to cross the membrane in a species-specific manner has been extensively studied in hope to use the acquired knowledge to develop new antimicrobials. It is known that using the unstructured N-terminus and receptor domain, colicins contact common porins at the outer membrane such as OmpF and BtuB and components of Tol and Ton machinery in the periplasm assembling a translocon. Although several crystal structures of colicin fragments with effector proteins (BtuB, TolB, OmpF) have been published, the architecture of the translocon remains unclear. To address this issue, we've studied colicin E9 in complex with its receptor and translocator proteins BtuB, OmpF and TolB by crosslinking mass spectrometry. As well as numerous internal cross-links within individual proteins, seven inter-protein crosslinks were identified. These data largely agree with the available crystal structures and previous MD simulations. However, some inter- and intra-molecular links cannot be explained with our current knowledge of the complex, so in this work I suggest four alternative models of the translocon structure. Clear functional separation between the domains makes colicins attractive candidates to use as the delivery chassis for other peptide and protein antimicrobials, which cannot otherwise cross the membrane or are not specific to pathogenic bacteria. To test whether this is feasible, I used the intrinsically disordered N-terminus of colicin N as a vehicle to deliver a BamD-binding peptide into the periplasm of *E. coli*. The delivery of the BamD epitope to the periplasm was successful, however the cytotoxic effect of this delivery was limited.

Acknowledgements

I would like to express my gratitude to Professor Colin Kleanthous who gave me an opportunity to work in his laboratory studying colicin translocation and introducing me to my collaborators Prof Shabaz Mohammed and Dr. Phil Stansfeld without whose expertise this work would not be the same. Professor Kleanthous has provided invaluable support and advice throughout this project to help me develop my scientific research and writing skills. I am deeply grateful to Prof Shabaz Mohammed for providing training in mass spectrometry theory, and for his helpful advice and discussion on a variety of topics: from data analysis and communication, to career development. I would also like to take the opportunity to thank Dr. Michael Cundell, and Svenja Hester for their help with proteomics wet lab methods, Dr. Katarina Jansen and Dr. Nicholas Housden for initial training in laboratory methods used in Kleanthous lab and helpful discussions, and Dr. Phil Stansfeld for his help with setting up and using the docking software.

I will be eternally greatful to my friends and family who have supported me throughout my studies. I would like to say a special thank you to my parents, Natalia and Vladislav Demyanenko, who have supported me through every step of the way and encouraged me to follow my dreams, keep my head up and aim high no matter what.

Writing this thesis was everything: interesting, extremely challenging, stressful, exciting, and sometimes frustrating, so my final acknowledgment is to my boyfriend Connor Sharp, who, despite having his own doctoral thesis, several papers and a patent to worry about, has found the time and energy to support me throughout the year, especially during the last few months. Thank you for keeping me healthy and sane, for listening to my babbles about this thesis 24/7, and for helping me put all ‘the’s and ‘a’s in the right places.

Table of Contents

ABSTRACT **2**

ACKNOWLEDGEMENTS **3**

ABBREVIATIONS **8**

1. INTRODUCTION **9**

1.1 CLASSIFICATION OF COLICINS AND THEIR TARGETS 9

1.2 COLICINOGENIC PLASMIDS 11

1.3 REGULATION OF COLICIN SYNTHESIS AND RELEASE 11

1.3.1 Organisation of colicin operon *12*

1.3.2 Synthesis of colicins, immunity and lysis proteins *13*

1.4 MODULAR STRUCTURE OF COLICINS 15

1.4.1 Colicin E9 ..... *17*

1.4.2 Colicin N *18*

1.5 RECOGNITION OF COLICINS AT THE SURFACE OF THE TARGET CELL 19

1.5.1 Receptor interactions *20*

1.5.2 Translocon formation at the OM *22*

1.5.3 Current model of ColE9 translocon formation *24*

1.6 TRANSLOCATION OF COLICINS ACROSS THE OUTER AND INNER MEMBRANES 25

1.6.1 Translocation of group A colicins *25*

1.6.2 Translocation of group B colicins *26*

1.5.5 Translocation of nuclease colicins across the IM *27*

1.7 MECHANISMS OF COLICIN CYTOTOXICITY AND IMMUNITY 28

1.7.1 Cytotoxic activity of pore-forming colicins and immunity *28*

1.7.2 Cytotoxic activity of enzymatic colicins and immunity. *29*

1.8 POTENTIAL USAGE OF COLICINS AND OTHER BACTERIOCINS AS ANTIBACTERIAL MOLECULES 32

1.9 POTENTIAL ISSUES WITH USING CLASS III BACTERIOCINS AS ANTIMICROBIALS 34

1.9.1 Intrinsic non-specific and temporary resistance *35*

1.9.2 Acquired specific resistance *35*

1.9.3 Other issues *36*

1.10 OTHER APPLICATIONS OF BACTERIOCIN BIOLOGY 37

AIMS OF THIS WORK **38**

OBJECTIVES **39**

2. MATERIALS AND METHODS **40**

2.1 STRAINS, MEDIA AND GROWTH CONDITIONS 40

2.2 GENERATION OF MUTANT COLICIN N PLASMIDS 42

2.2.1 Primer design *42*

2.2.2 Polymerase chain reaction (PCR) *43*

2.2.3 Competent cell preparation *44*

2.2.4 Transformation of chemically competent cells *44*

2.2.5 Purification of plasmid DNA *45*

2.3 PROTEIN EXPRESSION AND PURIFICATION 45

2.3.1 Ni-IMAC *45*

2.3.2 Size exclusion chromatography *46*

2.3.3 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) *47*

2.3.4 Protein concentration determination *47*

2.4 COLICIN ACTIVITY TEST 48

2.5 LYSINE-SPECIFIC CROSS-LINKING OF COLE9 TRANSLOCON 48

2.5.1 Cross-linking reaction *49*

2.5.2 2-D Blue-Native PAGE *49*

2.5.3 In-gel trypsin digestion *49*

2.6 LC-MS/MS 50

2.7 MASS SPECTROMETRY DATA ANALYSIS 50

2.8 PROTEIN-PROTEIN DOCKING AND MODEL ANALYSIS 51

3. COLICIN N AS A CHASSIS FOR TOXIC BAMD-BINDING PEPTIDE 52

3.1 INTRODUCTION 52

3.1.1 Outer membrane of Gram-negative bacteria 52

3.1.2 The BAM complex 53

3.1.3 BamA C-terminal peptide binds to BamD and inhibits growth of E.coli 54

3.1.4 Experimental design 54

3.2 RESULTS 56

3.2.1 Test expression of Colicin N-SAGIALQW constructs 56

3.2.2 Full-scale expression and purification of colicin N-SAGIALQW constructs 58

3.2.3 Effect of the SAGIALQW peptide insertion on activity of full-length colicin N 58

3.2.4 Suitability of JM83 and B^E3000 cells for assessment of colicin $N_{\Delta 186\text{-}387}$ 61

3.2.5 Effect of different media on colicin $N_{\Delta 186\text{-}387}$ toxicity in B^E3000 cells 62

3.2.6 Effect of the peptide insertion on the activity of truncated colicin $N_{\Delta 186\text{-}387}$ in BE3000 cells 63

3.3 DISCUSSION 65

3.4 FURTHER RESEARCH QUESTIONS ON COLICIN N AND *E. COLI* CELL BIOLOGY 65

4. PROBING THE INTERFACE BETWEEN COLICIN E9 AND ITS TRANSLOCON PROTEINS BY CROSSLINKING MASS SPECTROMETRY 66

4.1 INTRODUCTION 66

4.1.1 Chemical crosslinking for probing protein-protein interactions 68

4.1.2 Lysine-specific crosslinking 69

4.1.3 Unambiguous identification of crosslinked peptides and residues using tandem mass spectrometry 70

4.1.4 Experimental design 72

4.2 RESULTS 74

4.2.1 Optimisation of the lysine-specific crosslinking protocol for purified translocon 74

4.2.2 Crosslinking the translocon with BS2 and BS3 crosslinkers 76

4.2.3 Identification and validation of crosslinks *77*

4.2.4 Probing the flexibility of colicin E9 outside the translocon *81*

4.2.5 Changes in conformation of colicin E9 when it is part of the translocon *83*

4.2.6 Mapping intermolecular crosslinks to the current translocon *86*

4.2.7 Remodeling the translocon complex to fit the crosslinking data *88*

4.2.8 New model for the colicin E9 translocon *92*

4.3 DISCUSSION 95

4.3.1 The impact of these results on current understanding of ColE9 recognition and translocation *95*

4.3.2 Possible future work *97*

5. MAJOR CONCLUSIONS AND IMPACT OF THESE STUDIES ON FURTHER RESEARCH .. 100

5.1 USING COLICINS TO DELIVER TOXIC PEPTIDES INTO THE PERIPLASM 100

5.2 MAPPING THE TOPOLOGY OF THE TRANSLOCON 101

APPENDIX 103

REFERENCES 108

Abbreviations

ACN - Acetonitrile

BAM - β-Barrel membrane protein assembly

BN PAGE – Blue Native polyacrylamide gel electrophoresis

BS3, BS2 – Bis(sulfosuccinimidyl) suberate, Bis(sulfosuccinimidyl) glutarate

DMSO – Dimethyl sulfoxide

FDR – False discovery rate

FRET – Förster resonance energy transfer

HPLC – High performance liquid chromatography

IPTG – Isopropyl β-D-1-thiogalactopyranoside

LC-MS/MS – Liquid chromatography

LPS – Lipopolysaccharide

MD – Molecular dynamics

MS – Mass spectrometry

NHS – N-hydroxysuccinimide

OBS – Outer membrane protein binding site

OM, OMP – Outer membrane, Outer membrane protein

PMF – Proton motive force

PMSF – Phenylmethane sulfonyl fluoride

POTRA – Polypeptide transport associated domain

PSM – Peptide spectrum match

SAM – Sorting and assembly machinery

SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis

TCEP - Tris(2-carboxyethyl)phosphine

TEAB – Triethyl ammonium bicarbonate

1. Introduction

Bacteria often use protein toxins targeting closely related species to gain competitive advantage. Members of this family of proteins were first described in 1925 by André Gratia and called colicins by the name of the species, which they were found to kill (*E. coli)*[1,2]. Since then many more were discovered in various species and along with toxic peptides from gram-positive bacteria were given the collective name bacteriocins. They are now classified into three classes: class I – small post-translationally modified peptides (<5 kDa), class II – small heat stable unmodified proteins (<10 kDa), class III – large heat unstable proteins. The work described in this thesis focuses on class III bacteriocins from *E.coli* - colicins. They are highly efficient and species specific toxins, which do not need secretion systems to be exported, have modular structure, parasitise generic cellular machinery of the target cells to translocate, and are synthesised together with specific immunity proteins to protect producer strains. Some of these qualities like high efficiency, specificity and modularity make colicins an attractive alternative to traditional antimicrobials. This chapter will provide a general overview of colicins, their diversity, structure and function, as well as discuss their possible future applications.

1.1 Classification of colicins and their targets

The classification of colicins was proposed by Davies and Reeves in 1975 when it was discovered that mutations in either *tolA* and *tolB* or *tonB* proteins rendered bacteria tolerant to some colicins but not others[3,4]. Thus colicins which required TolA/B proteins for cytotoxic activity were called group A colicins, and those which required TonB - group B. Colicins of either group can be further classified by the type of their cytotoxic domain: enzymatic or pore-forming. Enzymatic colicins are usually nucleases (RNases or DNases), however there are exceptions like colicin M, the cytotoxic domain of which degrades lipid I

and lipid II (peptidoglycan precursors). Colicins with classification by group, function of the cytotoxic domain as well as their known receptor and translocator proteins are listed in table 1.1[5,6].

Table 1.1 Colicin classification by group, cytotoxic domain and required proteins for activity

Colicin name	Group	Cytotoxic domain type	Receptor proteins	Importer proteins
A	A	Pore	BtuB	OmpF, TolABQR
B	B	Pore	FepA	TonB, ExbBD
D	B	tRNase	FepA	TonB, ExbBD
E1	A	Pore forming	BtuB	TolC, TolAR[6]
E2	A	DNase	BtuB	OmpF, TolABQR
E3	A	rRNase	BtuB	OmpF, TolABQR
E4	A	rRNase	BtuB	OmpF, TolABQR
E5	A	tRNase	BtuB	OmpF, TolABQR
E6	A	rRNase	BtuB	OmpF, TolABQR
E7	A	DNase	BtuB	OmpF, TolABQR
E8	A	DNase	BtuB	OmpF, TolABQR
E9	A	DNase	BtuB	OmpF, TolABQR
Ia	B	Pore	Cir	TonB, ExbBD
Ib	B	Pore	Cir	TonB, ExbBD
K	A	Pore	Tsx	OmpF, OmpA, TolABQR
M	B	Lipid I and lipid II degradase	FhuA	TonB, ExbBD
N	A	Pore	OmpF/LPS	OmpF, TolAQR
S4	A	Pore	OmpW	OmpF, TolABQR
U	A	Pore	OmpA	OmpF, LPS, TolABQR
28b	A	Pore	OmpA	OmpF, LPS, TolABQR
5	B	Pore	Tsx	TolC, TonB, ExbBD
10	B	Pore	Tsx	TolC, TonB, ExbBD

1.2 Colicinogenic plasmids

Colicins are synthesised by many strains of *E.coli* including those found in healthy human gut and those associated with infections[7,8]. Producer strains carry genetic information for production of a colicin, its immunity and release proteins on a colicinogenic plasmid[9]. These plasmids can be of two types: type I plasimds which typically encode group A colicins, or type II which in turn encode group B colicins[10]. Type I colicinogenic plasmids are small (up to 10kb) and are present in relatively high copy numbers at around 20 per cell. On the other hand type II plasmids are large and may contain several colicin genes and are normally present in cells only one at a time. Both, type I and type II plasmids can be shared with other cells, however to be transmitted type I plasmids require the presence of a type two colicin plasmid or another conjugative plasmid. It is worth noting that replication of these plasmids is independent of colicin induction[11,12].

1.3 Regulation of colicin synthesis and release

The production of colicins is induced during the global response to DNA damage known as the SOS response[5]. Several different conditions which induce colicin production have been identified: UV light, ompR mutation, stationary growth phase, high temperatures, general nutrient depletion, thymine starvation and more. In the laboratory, the most common way of inducing colicin production is introduction of the DNA damage inducing antibiotic mitomycin C. Once colicins are synthesised they are released from cells by the action of specific colicin release proteins which initiate quasilysis, alter cell envelope structure and activate outer membrane phospholipase A[5]. Colicin synthesis, except for those encoded on type II plasmids, is fatal to producer cells, therefore they need to be well regulated.

1.3.1 Organisation of colicin operon

Genetic organisation of group A colicins is different to that of group B. These differences have been reviewed comprehensively by Riley[13,14] and are summarised below and shown in figure 1.3.1. The organisation of the colicin operon depends on the type of plasmid which carries the colicin gene, and on whether said colicin is an enzyme or a pore former with the exception of colicin M. Regardless of the plasmid type and mechanism of action, all colicin operons start with the gene for the colicin itself under the regulation of the LexA promoter. If the encoded colicin is a nuclease, then the colicin gene also contains a constitutive promoter for its immunity protein which is located directly downstream of the toxin gene. On the other hand, if the colicin is a pore former, the immunity gene would be located on the opposite strand to the colicin between the colicin release gene and the main colicin gene[15–17]. Although colicin M is an enzymatic colicin, its immunity gene is located on the opposite strand to the colicin gene and is expressed separately, like that of pore formers[18]. The genes for colicin and immunity proteins can also be followed by the colicin release protein gene. Specifically, this gene is found in all colicins encoded on type I plasmids[14], which includes all group A colicins and sometimes colicins 5 and 10 from group B[5]. Unlike immunity proteins, release proteins are often quite similar and can be used interchangeably by different colicins. Although each colicin usually has distinct immunity and release genes, there is some redundancy. Sometimes colicin operons contain extra genes: immunity genes for other colicins, and other lysis genes (type I plasmids) or a second colicin gene (type II plasmids)[5].

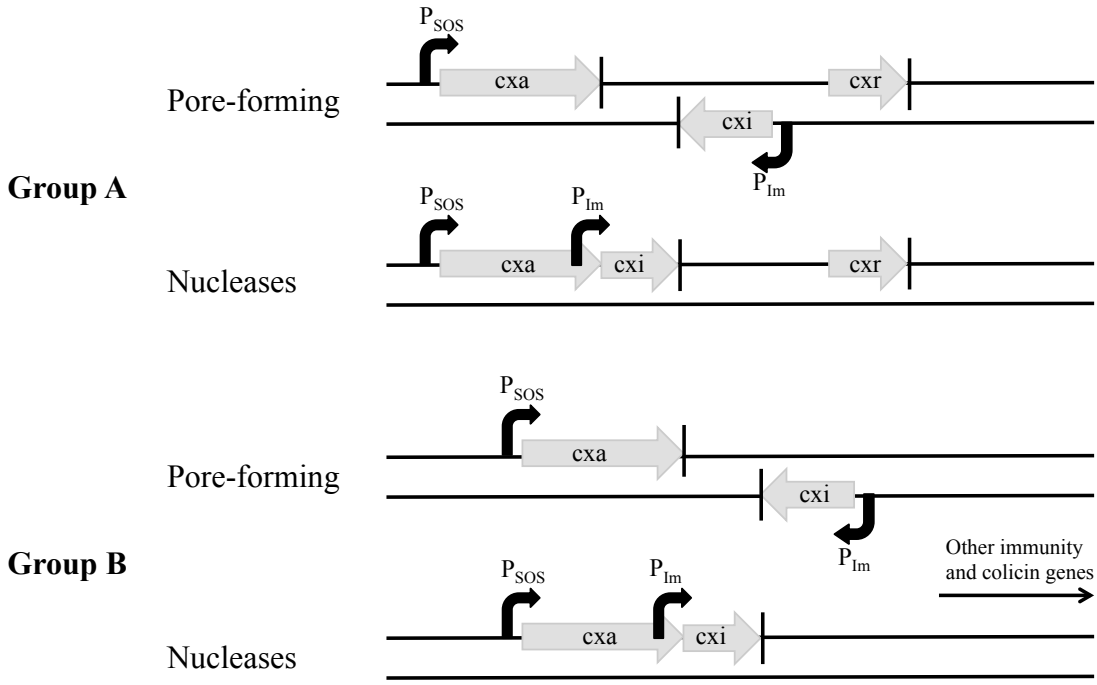


Figure 1.3.1 Organisation of colicin plasmids. The genes are shown as arrows and labelled according to a three letter convention: *c* stands for colicin, *x* for the letter denoting the variable name of the colicin, *a* stands for activity gene, *i* for immunity, and *r* for release. Promoters are shown as arrows and terminators as a straight line through the strands.

1.3.2 Synthesis of colicins, immunity and lysis proteins

While immunity proteins are constitutively expressed through their own promoters, the transcription of the entire operon including a colicin and a colicin release gene is regulated by LexA[19], although other proteins such as cyclic AMP receptor proteins have been shown to play a part in regulation of colicin transcription as well[20,21]. For most colicins, in normal conditions the LexA protein dimers are bound to two overlapping LexA binding sites (LexA boxes) within the operon. When DNA damage occurs, the RecA pathway is activated and promotes autocleavage of LexA and its release from the LexA boxes, releaving a DNA bend caused by LexA binding and allowing initiation of transcription. As a result, colicin as well as the release protein start to accumulate inside the cell. Once the concentration of the colicin release protein reaches the threshold level, synthesised colicins are released from the cell. This process was reviewed by Cascales and is summarised in

figure 1.3.2. Since colicin expression is lethal to the cells, producers of some colicins are known to employ safety net mechanims: colicin K expression is also regulated by a second repressor IscR, while expression of colicin E8 is co-regulated by AsnC. These proteins stabilise LexA preventing its cleavage induced by DNA damage, providing an additional regulatory step to ensure colicin is synthesised and released only as a last resort[22,23]. Colicin release can also be regulated by enviromental cues. In fact, suppressors of colicin release such as CsrA were found to play a role in supporting biodiversity in the bacterial community[24].

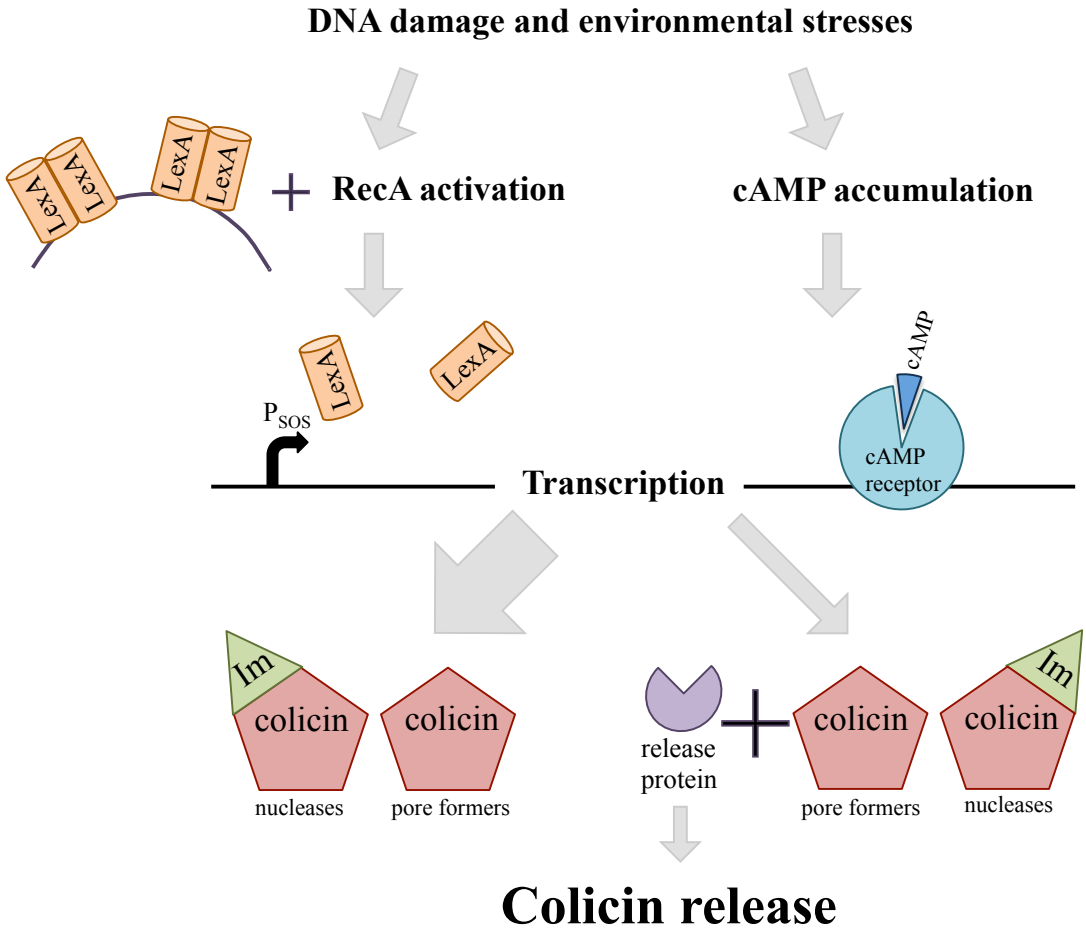


Figure 1.3.2. Colicin synthesis and release mechanism. In this diagram, LexA acts as the transcription repressor and cAMP receptor as activator. The larger arrow after transcription represents a major mRNA product, which only encodes the toxin and, in case of nuclease colicins, an immunity protein. The thinner arrow after transcription indicates the minor mRNA product, which in addition to colicin and immunity protein encodes the colicin release protein, accumulation of which leads to quasilysis and colicin release.

1.4 Modular structure of colicins

Colicins are elongated proteins[25,26] with molecular range between 30-90 kDa[25,27–31] and are comprised of three domains. The often disordered N-terminal domain which interacts with the host proteins to aid translocation across the membrane (T domain), followed by the central domain which interacts with the receptor protein (R domain) and finally the C-terminal cytotoxic domain (C domain) as shown in figure 1.4 A. The N-terminus of colicins is typically intrinsically disordered for up to 90 residues and contains regions that bind to translocator proteins with high affinity[5,32–37]. It is usually connected to the R domain via a globular region. The R domain is mostly alpha helical, and in colicins E3, E9 and Ia it is a coiled coil, although in colicin Ia it also contains a group of two anti-parallel β-strands and two small helices[38–40]. The C-terminal domains of pore forming colicins are α-helical, those of DNases are partly α-helical but include a functionally important β-sheet, while the structure of cytotoxic domains of colicin M and RNases are predominantly comprised of β-sheets[41–43].

DNase domains have very similar three-dimentional structures and high sequence identity (≥65%) as shown in figure 1.4 B. They usually consist of an α-helical shell, which surrounds the three-stranded anti-parallel β-sheet. The acive site of the DNase domain is located at the C-terminus and is part of the ββα-Me^{2+} structural motif present in a wide group of nucleases[44]. It contains a functional motif which is highly conserved in DNase colicins and also found in other endonucleases[45]. This motif was first described by Shub et al in 1994 and dubbed the H-N-H motif by the three residues which play the key roles in DNase activity[46,47]. The first His residue is largely responsible for the DNA hydrolysis, the middle Asn residue is vital for structure stability, and the final His residue is involved in binding divalent metal ions such as Mg^{2+}, Ca^{2+}, Zn^{2+}, Ni^{2+}.

RNase domains are diverse and have little structural similarity to one another, albeit they share some characteristics of RNA binding proteins. They are predominintly organised into β-sheets, with peripheral α-helices (fig 1.4 B).

Finally, pore forming colicins have similar tertiary structures as shown in figure 1.4 B. They consist of 10-helix bundle: 8 α-helices surrounding a central hairpin formed by two hydrophobic helices. Overall, the pore forming domain is a compact soluble protein, which upon structural rearrangements becomes a helical inner membrane pore.

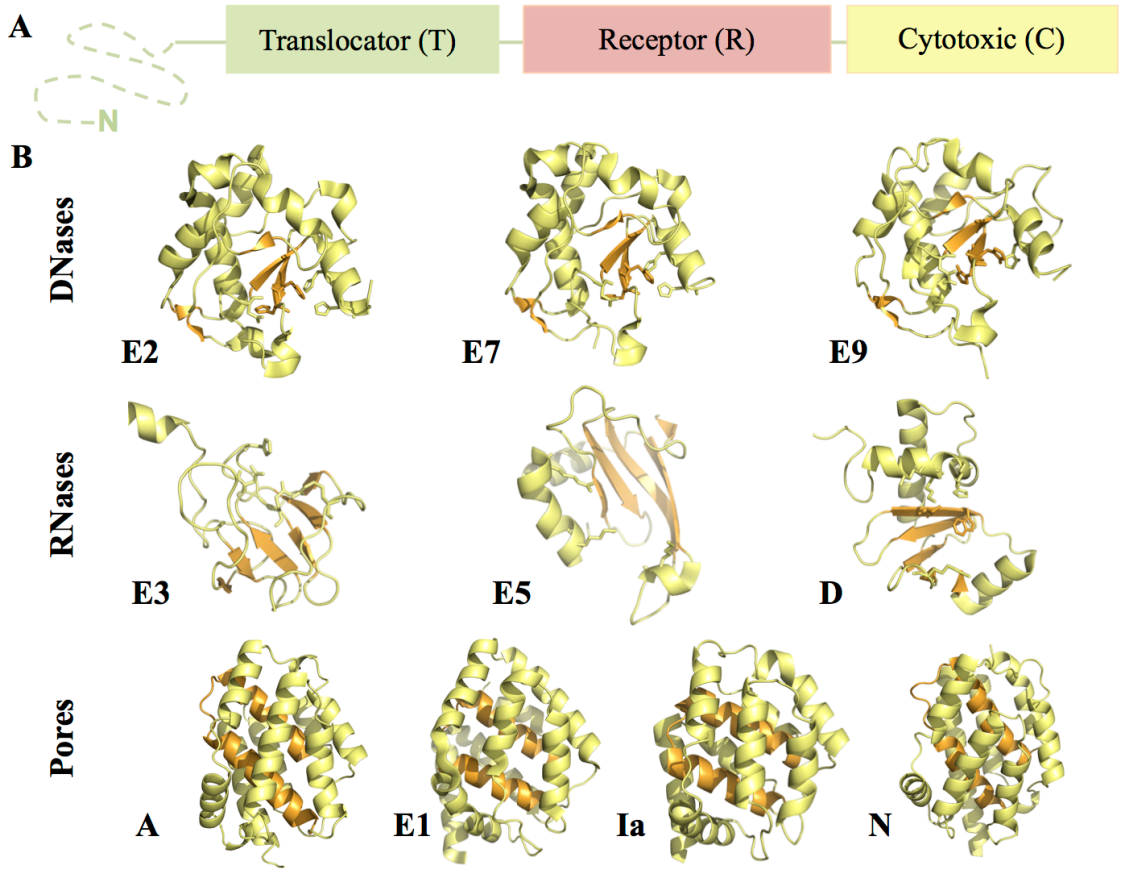


Figure 1.4 A. Organisation of colicins including the disordered N-terminus shown in dashed line as part of the translocation domain, followed by the receptor domain and C-terminal cytotoxic domain. **B**. Structures of C-terminal cytotoxic domains with beta sheets highlighted in orange and important residues shown in stick model for DNases and RNases. For pore-forming domains, the hydrophobic helices are coloured orange. PDB IDs left-to-right, top-to-bottom: 3U43, 1PT3, 5EW5, 1JCH, 2DFX, 1TFO, 1COL, 2I88, 1CII, and 1A87

Although overall homology between colicins is low, several colicins have the same T and R domains but different cytotoxic domains; for example, E3(rRNase) and E9 (DNase). The following sections will elaborate on the structure and function relationships in the DNase colicin E9 and the pore fore forming colicin N.

1.4.1 Colicin E9

Colicin E9 has three distinct domains. The N-terminal domain which interacts with OmpF and TolB[33], followed by the receptor domain which interacts with BtuB[48], and the cytotoxic H-N-H DNase domain[49]. The crystal structure of colicin E9 was published recently by Klein et al.[38] and is shown in figure 1.4.1. It is highly similar to the crystal structure of colicin E3 previously determined by Soelaiman et al.[39] which has the same T-R domains (>95% sequence identity), but an rRNase cytotoxic domain. The first 84 residues of the N-terminus are missing from the electron density, which suggests that this part of the protein is disordered. The function of this sequence in Colicin E9 is well mapped and contains two OmpF binding sites termed OBS1 (residues 2-18) and OBS2 (residues 54-63), and a TolB binding site (TolB-box) between them (residues 32-47)[50,51]. The disordered N-terminus is followed by 232 residues which are structured into a mixture of α-helices and curved beta-sheets, the function of this sequence remains unclear. It is followed by the coiled-coil of the R-domain, which is known to contact the BtuB protein through residues 369-402[48]. Finally, the alpha helical H-N-H/ββα-Me^{2+} DNase domain is located at residues 453-580. As colicin E9 is a nuclease colicin, it is synthesised together with its immunity proten, which is tightly associated with the DNase domain after synthesis, but lost during the translocation process[52,53].

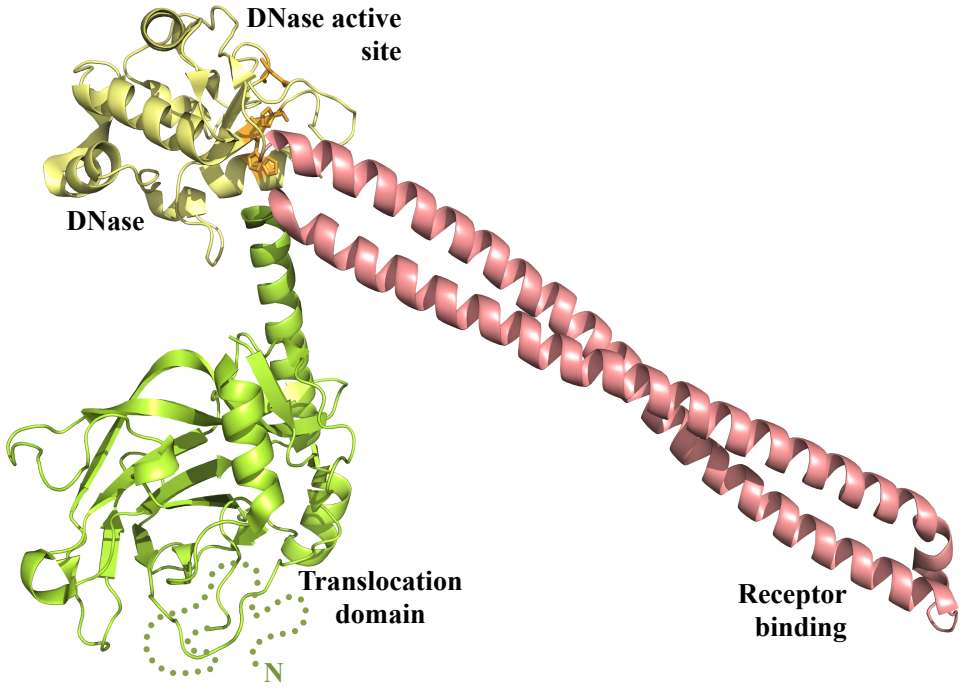


Figure 1.4.1 Crystal structure of colicin E9 (PDB ID: 5EW5). The translocation domain with the disordered N-terminus shown with dotted line, receptor-binding domain in coral and the DNase domain in yellow, with the crucial residues involved in the nuclease activity are shown in orange.

1.4.2 Colicin N

Colicin N is a pore forming colicin which contacts only one protein in the outer membrane of *E. coli* – OmpF[54]. To translocate across the OM colicin N also requires inner membrane proteins TolAQR to be present. Although it has all of the main features of colicins, the domain organisation of colicin N, unlike that of colicin E9 is not well understood. Its crystal structure was solved by Vetter in 1998 and is shown in figure 1.4.2[55]. The first 90 residues are not visible in the density plot as they are intrinsically disordered. However, it is known that this part of the protein contains at least one OmpF binding site OBS1 at the extreme N-terminus (approx. residues 2-15[56]), and TolA binding site at residues 43-70[57]. The function of other regions of the disordered T domain is not clear. Unlike colicin E9, the R-domain of colicin N (91-183) has six β-strands wrapped around an α-helix, which becomes part of the pore forming domain towards the C-terminus. It has been suggested

that the R-domain of colicin N selectively binds to lipopolysaccharide (LPS) and acts as a primary receptor for LPS near the OmpF trimers[58]. The pore forming domain is located at the C-terminus of the protein (residues 184-387) and like other pore domains is composed of a 10-helix bundle, which includes the α-helix of the R-domain.

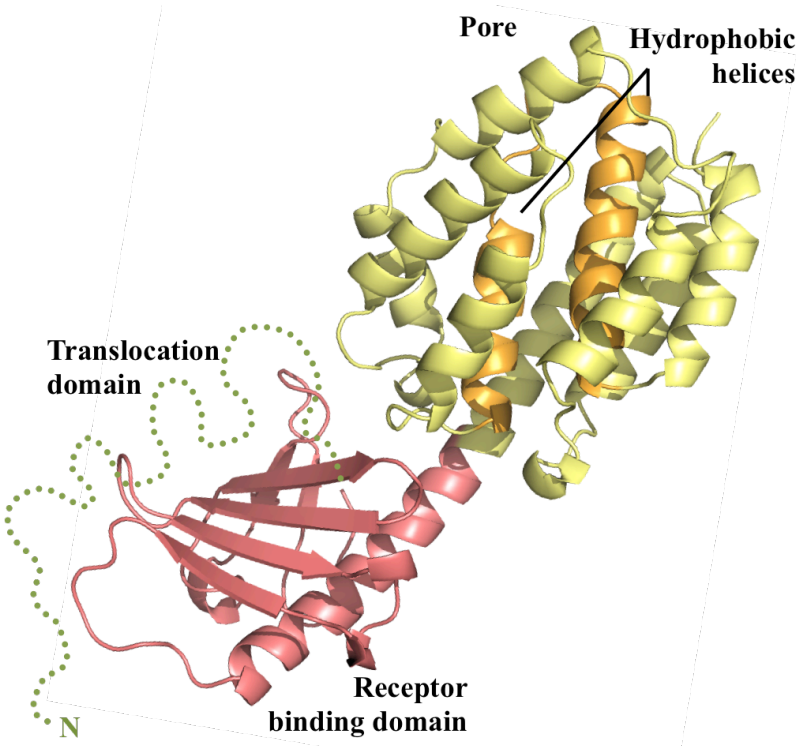


Figure 1.4.2 Crystal structure of colicin N. The disordered 90 residues N-terminal translocation domain, which is absent in the structure, is shown with the green dotted line (PDB ID: 1A87). The β-sheet is shown in coral as part of the receptor-binding domain. The C-terminus is shown in yellow with the hydrophobic helices in orange.

1.5 Recognition of colicins at the surface of the target cell

After release from the producer cells, colicins need to be imported into the target cells. The mechanism of colicin import includes contacting their primary receptors at the host cell outer membrane, locating translocator proteins to start the process of translocation, and contacting the proteins which provide the energy for translocation. The mechanisms of these processes vary depending on the group to which the colicins belong to, i.e whether the colicin requires TolABQR or TonB-ExbBD system. The mechanisms of translocation

are complex and vary from colicin to colicin. Although the mechanisms involved are not fully understood, the following sections provide an overview of current knowledge of colicin translocation mechanisms.

1.5.1 Receptor interactions

Group A and B colicins use different outer membrane proteins as receptors and translocators. As a rule, group A colicins, use OmpF (most group E colicins can also use OmpC, PhoE[59]) or TolC as their translocator. Unlike group A colicins, those in group B require tonB-dependent transporters like Cir, FhuA and FepA. Although there are exceptions: colicins 5 and 10, which require Tsx and TolC for their translocation. For some group B colicins it has been shown that while they only require one type of protein, they may need two copies of them to use, one as a receptor and one as a translocator protein (e.g colicin Ia)[37].

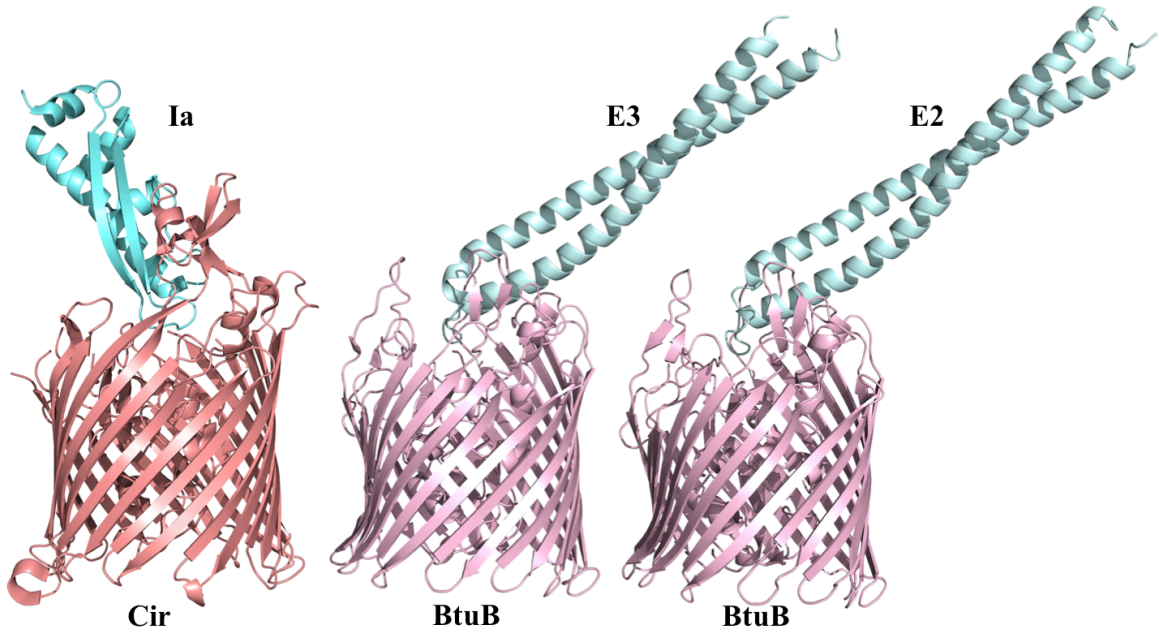


Figure 1.5.1 Receptor binding domains of colicins Ia, E3, and E2 bound to their receptor proteins Cir and BtuB (PDB Ids: 2HDI, 1UJW, 2YSU). The orientation of Cir relative to BtuB was determined by structural alignment of the two barrels. At this position the angle between receptor-binding domains is close to 180°. Since the receptor binding domains of colicins E3 and E2 have extremely high sequence identity they are shown in the same shade, as their receptor proteins BtuB.

Most receptors for both group A and group B colicins (with the exception of colicins N, U 28b), are common TonB-dependent transporters such as BtuB, Cir and FhuA. Despite having low sequence identity <30%, these proteins have similar tertiary structures: 22-stranded β-barrels with N-terminal plug domains and long flexible extracellular loops which are stabilised by metal and ligand binding[60,61]. Colicins bind their receptors at the apex of the R domain with high affinity with average dissociation constant (K_d) estimated around 0.1-1 nM[48,50,62]. Residues of the R domain contact several residues of the receptor plug domain, however most of the colicin-receptor contacts are made to the flexible loop regions[48,63,64]. Structures of three R domains with their receptor proteins have been published and are shown in figure 1.5.1[48,63,65]. How colicins B, D and M interact with their receptor and translocator proteins is yet unknown.

As shown in figure 1.5.1 interactions of the R domains of colicin Ia and E3 with their receptors Cir and BtuB are similar in nature, with the R domain at approximately 45° relative to the membrane bilayer, no apparent change to the plug domain organisation and approximately 990-1500 $Å^2$ of buried surface area between the colicin and its receptor. However, there are a few key differences. In colicin E3-BtuB interaction, two residues R388 and Q398 in the R domain contact three different loops of BtuB and represent ~40% of all interactions, while 8 out of 11 extracellular loops are involved in binding colicin E3. While in colicin Ia-Cir interaction, 9 different residues of colicin Ia contact only two arginines (in two different extraxcellular loops of Cir) and are responsible for >50% of all interactions, while only 4 out of 11 extracellular loops are involved in binding colicin Ia. Another major difference between the two complexes is the large conformational change in extracellular loops of Cir induced by colicin Ia binding[63], which is not apparent in colicin E3-BtuB structure. Additionally, as seen from figure 1.5.1, the direction of colicin Ia binding to Cir is differerent to that of colicin E3 to BtuB by approximately 180°. The reason for this difference is unknown.

The interaction between colicins and their receptor proteins typically mimics the interaction of the physiological substrates with respective transporters, leading to competition between colicins and physiological ligands for the same receptors[48,66]. TonB-dependent transporters have higher affinity for their natural substrates than for colicins, therefore in most cases substrates can rescue colicin treated bacteria from colicin-mediated killing[5]. Except in the case of colicin B: despite the higher affinity of FepA for ferric enterobactin (natural substrate), its addition does not have a rescuing effect on colicin B treated cells[67].

1.5.2 Translocon formation at the OM

After being recognised by their receptor proteins, colicins recruit their translocators at the outer membrane along with proteins in the inner membrane and periplasm while they are still bound to the OM receptor, forming a translocon[33]. Its exact mechanism of formation for various colicins has long been a subject for debate, although the first stages, at least for colicins of group A, are well characterised. As mentioned above, most group A colicins use OmpF for translocation with the exception of colicin E1. The mechanism of OmpF and Tol machinery recruitment was recently discovered for colicin E9, and is described in the next section and summarised in figure 1.5.3[33]. This mechanism could potentially be applicable to all colicins of group A which use OmpF and TolABQR as their translocator proteins.

Unfortunately, for colicin E1 and N from group A, the mechanism of translocon formation is less clear. Colicin E1 uses TolC for translocation, which is an outer membrane channel frequently implicated in multi-drug resistance, acting in combination with AcrAB as a multi-drug efflux pump[68,69]. Although no structure of colicin E1 has been published, except for the pore forming domain, it is known that its R domain is similar to that of colicins E2 and E3, while its T domain (1-190) is known to be α-helical and contains a short

intrinsically disordered region (1-40) and no more than two α-helices in the rest of the domain[70]. The T domain of colicin E1 is thought to contact TolC at the apex of its helical hairpin close to the 3.4 Å constriction of the TolC, leaving the disordered N-terminus outside and no obvious route for translocation[70]. While colicin N is thought to only require OmpF for translocation, which has enough room for disordered T domain to thread through, due to its intercaction with LPS and conflicting evidence for insertion of the disordered region into the OmpF lumen, the mechanism of its translocation remains a mystery[71,72].

Translocon formation for group B colicins is even less well-understood. With the exception of colicins 5 and 10, group B colicins use only one kind of protein as a receptor and translocator. Although colicin Ia is known to use two copies of Cir, it is not known whether colicins B, D and M use one or two copies of FepA and FhuA. The most studied of these is colocin Ia and its interaction with CirA. Binding of the R-domain to the first copy of Cir leads to recruitment of the second copy of Cir which in turn interacts with the N-terminal domain of colicin Ia[37]. The disordered region of colicin Ia is thought to mimic the binding of the natural ligand and therefore would dislodge the TonB-box of the plug domain of Cir so that it becomes exposed to the periplasm, allowing its interaction with TonB and subsequent removal of the plug domain[37]. This would clear the way for the N-terminus of colicin Ia to move through the opened channel to contact either the same or different TonB molecule to form a translocon.

1.5.3 Current model of ColE9 translocon formation

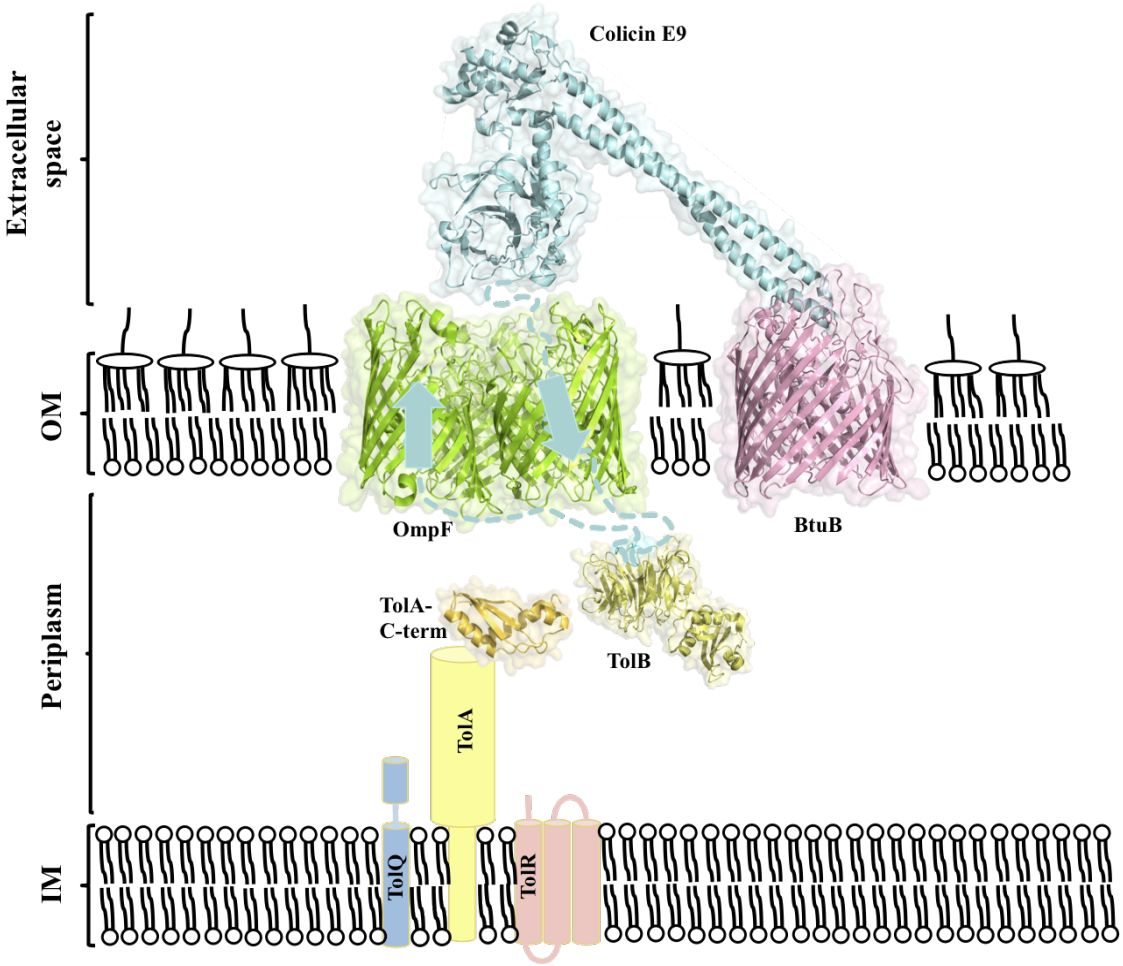


Figure 1.5.3 Current colicin E9 translocon model. Crystal structure of colicin E9 (PDB ID: 5EW5) was aligned to the crystal structure of BtuB to the colicin E3 receptor-binding domain (PDB ID: 1UJW), which has a virtually identical sequence to that of colicin E9). The OmpF trimer (PDB ID: 3O0E) was positioned in the membrane directly below the translocation domain of colicin E9. The interaction between OBS1 and OBS2 with OmpF is shown with dotted cyan line as these residues are missing in crystal structure but are known to bind inside the OmpF beta barrel in preferred orientation, which is indicated by arrows. Residues between OBS1 and OBS2 interact with TolB (PDB ID: 2IVZ). Crystal structure of C-terminus of TolA, which is known to interact with TolB is shown also (PDB ID: 1S62/1TOL). Other parts of TolAQR are shown as stylised models.

Translocon formation by colicin E9 has been studied extensively, our current knowledge is summarised in figure 1.5.3. Colicin E9 binds to its receptor BtuB with extremely high affinity (K_d 2.4nM-0.15μM)[50], then uses its disordered N-terminus to contact a nearby OmpF trimer. It threads through the lumen of the porin from the outside through to the periplasm, where the extreme N-terminus (OBS1) binds in the lumen of the neighbouring porin from the inside to the outside[33]. This exposes the TolB-box of colicin E9 in the periplasm and allows TolB to bind at its binding site which then leads to formation of a translocon complex at the outer membrane.

1.6 Translocation of colicins across the outer and inner membranes

After formation of the translocon, both group A and group B colicins require proton motive force (PMF) to be transported across the OM[26]. This energy is provided by TolABQR system for group A colicins and TonB-ExbBD for group B. It is assumed that another requirement for translocation is at least partial unfolding of the colicin[53,73,74]. Together with interactions with the Tol machinery it is thought to guide the Brownian ratchet mechanism for unidirectional transport of colicins across the OM, and has been shown to aid immunity protein release from the nuclease domains[74].

1.6.1 Translocation of group A colicins

Although the mechanisms of translocation for individual colicins within group A may vary, the general path is thought to be the same. The physiological role of Tol-Pal machinery is to maintain the stability of the outer membrane coupling to proton motive force (PMF), it also plays a part in cell-division[75,76]. TolQ and TolR are integral inner membrane proteins, the main functions of which are to activate periplasmic protein TolA that is anchored to the inner membrane via its N-terminus in PMF-dependent manner[77]. TolA then transfers this

energy to Pal, a periplasmic protein achored to the inner leaflet of the outer membrane, via a periplasmic protein TolB[78]. Group A colicins have evolved to hijack this system to enter bacterial cells. Except for colicins N and E1, group A colicins require TolB to connect the colicin at the periplasmic interface of OM with TolA[57,79], which in turn transfers the energy necessary for translocation. Colicins N and E1 contact TolA directly. Colicin A is known to interact with TolA as well, but unlike colicins E1 and N, it still requires TolB[80]. Although putative TolR boxes have been identified in several colicins, direct interaction of colicins with TolQR has yet to be demonstrated[5]. The distance between the OBS1 and TolAB binding regions of colicins is likely to be important for efficiency of their translocation. For instance, the distance between the OBS and TolA box in colicin N is longer than the distance between OBS and TolB box in colicins A, E2-E9 by approximately 10 amino acids. While this TolA box is also roughly in the same place, albeit slightly further down towards the C-terminus, as the TolA binding sequence of colicin A. This may be the reason for the extended disordered region of colicin N and lack of clearly definable regions other than OBS1 and Tol box compared to that of other colicins. The importance of this distance between OBS and TolA box of colicin N is also apparent in this study (see section 3.2.).

1.6.2 Translocation of group B colicins

TonB-ExbBD system is used by Gram-negative bacteria to facilitate import of small molecules: siderophores and cobalamin via TonB-dependent receptors (FhuA, FepA, BtuB) by providing energy for import[81]. As its function is similar to that of the Tol-Pal system, the organisation is also quite similar, but with fewer components. The function of ExbBD in this system is analogous to TolQR of Tol-Pal and is to harvest the energy generated by the PMF and transfer it, in this case, to TonB[82]. Thus while TolA and TonB are essential for translocation of respective colicin groups, partial complementarity with TolQR in the

absence of ExbBD and vice versa can be achieved[83–85]. Despite the functional similarities between TolA and TonB: anchored to IM by their N-termini, span the periplasm, transfer mechanical energy to protein in the OM, these proteins have low sequence identity at their C-termini. This is likely due to them playing different roles at the outer membrane and having different binding partners. Unlike TolA, TonB has a proline-rich hydrophilic periplasmic domain[86,87], which allows the protein to extend all the way to the OM to enable direct interaction between its C-terminus and integral OM proteins via a specific motif in their N-terminal plug domains called the TonB box[88–90]. Colicins of group B also have this motif in their N-terminal domains and use it to parasitize this system. However, the sequence of the motif is not conserved and subtle differences within TonB boxes have been shown to impact the mechanism of colicin translocation[91]. Consistent with the differences between TolA and TonB systems is the fact that compared to TolB and TolA boxes of group A colicins, TonB boxes of group B colicins are located much closer to their N-termini (residues 23-27 of colicin Ia, 16-20 of colicin B, 17-21 in colicin D, 2-6 in colicin M).

1.5.5 Translocation of nuclease colicins across the IM

Unlike pore forming and lipid I and II degrading colicins (N, A, B, M) nuclease colicins need to translocate across the inner membrane before reaching their target. It is known that due to their overall positive charge DNase domains are capable of forming short-lived voltage-dependent channels in the IM, where the IM AAA^+ ATPase/protease FtsH is involved in transporting the nuclease domains of both group A and B colicins to the cytoplasm[26,92]. Colicin D also requires the presence of a signal peptidase LepB for its activity, albeit the reason for this dependence is not known[92].

Although some nucleases have been shown to be cleaved by the OM associated protease OmpT, recent studies show that OmpT deletion mutants remain fully sensitive to colicins D and E3 and the cleavage of certain colicins may render their nucleases inactive therefore it is unlikely that OmpT is involved in this process[92,93]. It is possible that in this case, OmpT acts as a defence mechanism.

1.7 Mechanisms of colicin cytotoxicity and immunity

1.7.1 Cytotoxic activity of pore-forming colicins and immunity

As mentioned previously, pore-forming domains of colicins A, B, E1, N, Ia are α-helical and are essentially a hairpin made of two hydrophobic α-helices surrounded by eight amphipathic helices. This arrangement results in a domain soluble in aqueous environment, which upon contact with membrane phospholipids and some structural rearrangement can insert itself into the IM to form a voltage-dependent non-specific cation channel[94,95]. The mechanism of this insertion and the final structure of the pore in the IM are subject to discussion. Recent EPR and FRET experiments suggest that the closed channel resembles an "umbrella", in which the hydrophobic hairpin is inserted into the membrane like a handle and the amphipathic helical bundle forms a sort of canopy at the periplasmic face of the membrane[96–99]. It has also been shown that these structures are prone to oligomerisation (most likely dimerisation for colicin A pore-forming domain, 2-6-mer for colicin N)[100,101]. Upon application of positive voltage, a large cation pore with low conductance opens[102,103], this causes membrane depolarisation, ATP depletion and consequently cell death. The organisation of the open pore is unclear. However, it is known that not all helices participate in formation of the pore[104–106], which raises the issue of the pore being too large to be created by seven short helices. It is possible that pore formation is due to recruitment of phospholipids to form a channel, transient oligomerisation or interaction with other

amphypathic helices within the inner membrane[5,107]. This would be consistent with the proposed immunity mechanism where the immunity protein binds to the pore domain at the inner membrane to prevent pore formation[108,109].

1.7.2 Cytotoxic activity of enzymatic colicins and immunity.

Enzymatic colicins have four different types of cytotoxic domain: 16S RNases, tRNases, DNases, or lipid I/lipid II esterases. DNases catlyse DNA cleavage through the action of the H-N-H/ββα-Me motif. The mechanism of their action is analogous to the known mechanism of action of related DNases (I-PpoI, Vvn). Histidine residues bind to the phosphate backbone in the minor groove of the DNA molecule, which causes a localised bend and brings the phosphate group towards the divalent cation (Mg^{2+}, Ca^{2+}, etc.) as shown in figure 1.7.2A[110]. This strains the bond between the bridging oxygen and the rest of the phosphate group. At this point the cation is thought to stabilise the phosphoanion transition state and along with a histidine side chain activate a water molecule which will then protonate the 3′ leaving group at the cleavage site. The 5′ group is thought to be stabilised by an arginine residue in the active site.

Toxicity of rRNase colicins is caused by the cleavage of the 16S subunit of ribosomal RNA[111]. The rRNase of colicin E3 cleaves ribosomal RNA at residues A1493/1494G close to the decoding centre of ribosomal A site[112]. The crystal structure of C-terminal domain of colicin E3 bound to 70S ribosome was solved in 2010 and is shown in figure 1.7.2 C[112]. In this structure, colicin E3 cytotoxic domain is bound to 16S subunit in a similar way to nuclease RelE and release factor RF2: glutamine and serine residues at the active site bind to nucleobase 1492 pulling it out of the RNA helix, while an arginine and another glutamine residues staibilise the partially extruded nucleobase 1493[112]. Cleavage of the 16S

rRNA is thought to be accomplished through deprotonating action of a glutamate residue on the 2′-hydroxyl group of the A1493 leading to nucleophilic attack on the adjacent 3′-phosphate group, which results in formation of a transition state stabilised by a positively-charged histidine residue[112]. To complete the cleavage, the protonated group of the same histidine donates a proton to the 5′-hydroxyl group of G1494[112]. Catalytic mechanism of cytotoxic domains of colicins E4 and E6 are thought to be similar due to a high degree of sequence identity[5].

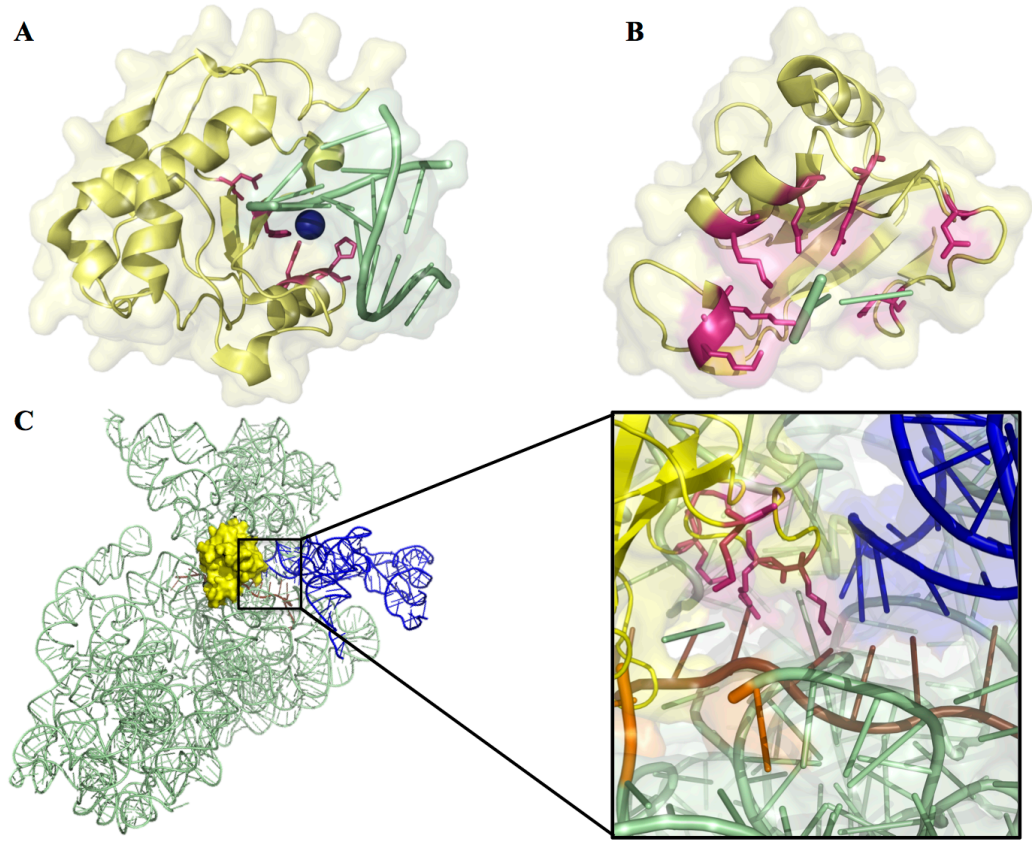


Figure 1.7.2 A. Crystal structure of the C-terminal DNase domain of colicin E9 in complex with Mg^{2+} ion and a DNA octamer. This is an inactive mutant H103A: still capable of DNA binding, but incapable of hydrolysis. **B.** Crystal structure of tRNase domain of colicin E5 in complex with a substrate analogue dGpdUp. Residues implicated in binding are highlighted in pink. **C.** Crystal structure of RNases domain of colicin E3 (yellow) in complex with 16S rRNA (pale green). For simplicity, ribosomal proteins and 50S subunit molecules were removed from the picture. To provide context in addition to 16S tRNA, mRNA is shown in brown, and a tRNA molecule at ribosomal P-site is shown in blue. Important residues on the colicin RNase are highlighted in pink.

Finally, different tRNase colicins are specific for different tRNAs: colicin E5 is known to cleave tRNAs which carry His, Asn, Tyr, and Asp, while colicin D cleaves four Arg isoaccepting tRNAs[113,114]. The mechanisms of action of these enzymes are thought to be similar to those seen in classic ribonucleases but with some differences in the topology of the active site which is shown in figure 1.7.2 B[41,42,115]. Mainly, unlike classic tRNase T1, colicin B active site contains no histidine residues, but instead has aspartate and arginine residues for catalysis[116]. Colicin D on the other hand, has only one histidine along with two stabilising lysine residues[42].

The immunity proteins for nuclease colicins are synthesised and released together with the toxin and protect the producer cells from their effects through tight binding to the enzymatic domain either at the active site (tRNases) or adjacent to it (DNases and rRNases)[26].

Colicin M is the smallest known colicin, and the only one which degrades peptidoglycan precursors by cleaving them into the lipid and pyrophosphoryl groups leading to reduced polymerisation and stability of the peptidoglycan layer[117]. Unlike other colicins, it requires an extra maturation step in the periplasm of the target cell by the periplasmic peptydyl-prolyl *cis-trans* isomerase FkpA, which is thought to catalyse the structural rearrangement of the colicin M active site[118]. The mechanism of action of the colicin M immunity protein is unknown, however it is known that it is not synthesised together with the toxin, but instead is targeted to the periplasm and anchored to the outer leaflet of the IM after synthesis[18].

1.8 Potential usage of colicins and other bacteriocins as antibacterial molecules

Antimicrobial resistance is a pressing issue in the treatment of nosocomial infections caused by bacteria sometimes referred to as the ESKAPE pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter, Pseudomonas aeruginosa, Enterobacter)*. Some clinical isolates are already resistant to all of the available antimicrobials on the market, and with no new antimicrobials, the proportion of these isolates will only grow with time. The need for new targets and approaches to antimicrobials is apparent.

Several members of the ESKAPE group, along with many other human pathogens are Gram-negative bacteria (*Klebsiella, Pseudomonas, Enterobacter, Neisseria),* which are inherently more difficult to treat due to the presence of the outer membrane with its lipopolysaccharide (LPS) coating. LPS is comprised of lipid A, a core oligosaccharide and a long chain polysaccharide called the O-antigen. It makes up the outer leaflet of the outer membrane, protecting the cell from enviromental stress by preventing diffusion of small hydrophobic molecules (for example detergents) into the cell and greatly reduces permeability of the membrane to common antibiotics[119]. The presence of LPS in the outer membrane is not the only defence mechanism of Gram-negative bacteria against harmful compounds such as antibiotics. There are several pathways through which Gram-negative bacteria can become drug resistant: reduced synthesis of porins through which the drugs can enter the cells, employing multidrug efflux pumps[120], alteration or degradation of antimicrobial agents[121], mutations in the target binding sites, and biofilm formation[122]. Therefore, to be successful, new antimicrobials need to be able to bypass these existing resistance routes and prevent development of new resistance mechanisms. Specificity of

antimicrobials to particular pathogenic species instead of broad-spectrum activity is also important to ensure that the native non-pathogenic bacterial communities are unharmed during the treatment to prevent dysbiosis. Moreover, targeting antibiotics to specific pathogenic species would remove the selective pressure from native bacteria, preventing accumulation of resistance genes to various antimicrobials, which can be passed on to pathogenic bacteria during infection. Specificity of bacteriocins towards a narrow range of species is one of the main advantages of using them as an alternative to conventional antimicrobial treatment.

Class I and class II bacteriocins have been used widely in the food industry and agriculture as a harmless alternative to traditional antibiotics. Bacteriocins from lactic acid bacteria are used successfully as preservatives either by addition of producer bacteria or purified compounds to food products[123]. Some of these bacteriocins have been shown also to be effective against bacterial spores which are typically resistant to antimicrobial treatments[124]. Several *in vitro* and *in vivo* studies were published showing significant effect of class I and class II bacteriocins against *C. difficile, S. aureus, S. pyogenes, S. pneumoneae, S. enterica*, *M. tuberculosis* and other human pathogens including their multi-drug resistant strains[125]. Although these bacteriocins are heat stable, amenable to bioengineering and highly effective, their applicability in the clinical setting may be hampered by complex design strategies and extensive post-translational modifications (for class I bacteriocins), development of resistance via traditional routes of drug-resistance development and immunity-mimicry, and proteolysis[125,126]. While some class I and class II bacteriocins like microcins are effective towards Gram-negative bacteria, the majority are directed at Gram-positive bacteria without dealing with the issue of multidrug-resistant Gram-negative pathogens.

Using class III bacteriocins in addition to class I and class II bacteriocins could help to resolve some of these issues. While also being highly specific to the species which they target, colicin-like molecules are large unmodified modular proteins which allows predictable and comparatively simple design of species specific toxins. Moreover, due to diversity of their cellular receptors and targets, development of resistance to combinations of class III bacteriocins due to receptor and target mutations is unlikely[127]. They are widespread across γ-proteobacteria and are known to be produced by known human pathogens such as extraintestinal pathogenic *E. coli* (ExPEC)*, Klebsiella sp., P. aeruginosa* and their relatives. Colicins and colicin-like bacteriocins have already been shown to be effective against biofilm associated bacteria *in vitro*[128], as a preventative measure against colonisation of urinary catheters[129], and prevention of post-weaning *E. coli* related diarrhea in pigs[130].

1.9 Potential issues with using class III bacteriocins as antimicrobials

Despite their effectivness and advantages over other antimicrobials, several issues with their applicability to a clinical setting may arise: strains may still become resistant to individual bacteriocins through either specific or non-specific resistance mechanisms. This is especially likely if producing strains are used instead of purified bacteriocins due to potential transfer of immunity genes to the target strain. In addition, the delivery of adequate concentration of active bacteriocins to the active site may be problematic due to their size. As well as resistance and delivery issues, economic factors related to production, purification, and storage may complicate further usage of class III bacteriocins, due to them being large proteins prone to inactivation by denaturation and proteolysis.

1.9.1 Intrinsic non-specific and temporary resistance

Resistance to colicins can be conferred upon otherwise susceptible cells by long chain, and S-type chain O-antigen (10-28 O-antigen repeat units), which masks colicin binding sites on the outer membrane porins regardless of the colicin type, while strains with short (2-14 O-antigen repeat units) and very long (>80 O-antigen repeat units) remain sensitive[131]. Length of the core oligosaccharide of the LPS may also be a resistance factor, as it was shown to be important in sensitivity to colicin N[132]. Moreover, it has been suggested that the environment in which bacteria grow can also have an effect on whether or not the given strains would be susceptible to colicins. For example, certain colicins preferentially target biofilm associated bacteria and are ineffective against planktonic cells of the same strain, while others preferentially target planktonic bacteria and are not active against biofilm-associated bacteria[128]. This is likely because certain growth conditions like nutrient content of media, temperature etc. may regulate expression of colicin receptors, the environment may also play a role in colicin resistance[133]. This kind of effect does not arise from changes/polymorphisms in cellular receptors and targets of particular colicins, but instead relies on intrinsic defence function and natural regulation of protein expression independent of colicin exposure.

1.9.2 Acquired specific resistance

Apart from non-specific resistance described above, bacteria can become resistant to specific colicins via natural selection by exposure to sub-minimal inhibitory concentrations (MIC) of the colicin in question. This pathway may include: accumulation of mutations in the receptor and translocator proteins, change in gene expression pattern in response to colicin treatment, and of course by acquisition of the immunity gene for that particular colicin. It has been shown that a point mutation in the lumen of the OmpF (G119D), which

results in constriction of the pore, can render colicins inactive against cells harbouring this mutation[134]. These mechanisms typically come at a fitness cost similar to those which confer resistance to tranditional antibiotics[135]. As most mutations would occur in the active sites of receptor and translocator proteins or result in truncated gene products, the native function of these proteins is likely to be affected. This type of resistance can be alleviated by using a combination of colicins which rely upon different outer membrane porins for translocation and target several aspects of bacterial physiology (i.e growth and replication (DNases and RNases), membrane and peptidoglycan intergrity (pore-formers and colicin M))[127]. Some bacteria can acquire, or already posess immunity genes for several class III bacteriocins produced by related species[136]. However, this issue may be circumvented by an engineering approach. High specificity of immunity proteins to particular nuclease domains[137], and the modular nature of bacteriocins can allow recombination of receptor and translocator domains from one species to be combined with a cytotxic domain from a bacteriocin produced by another genus, for which the target species would not have an immunity protein.

1.9.3 Other issues

Because of their specificity to particular bacteria and sensitivity to pepsin and trypsin activity, colicins are considered safe and non-toxic to mammalian or plant cells. However, because of this, colicins would face some additional issues when used as antimicrobial compounds, which need to be taken into account. For example, changes in the environment could affect the folding state of the colicins making them unable to contact receptors and translocators at the target cell surface. If colicins come into contact with non-specific proteases, they will be degraded to small peptides, which will not have the desired effect on target bacteria. Similarly, the need for colicins to interact with the OM of sensitive bacteria

would render them ineffective against intracellular pathogens. Although these issues do exist, it does not mean that bacteriocins could not and should not be used as antibacterial molecules. In fact, depending on the application, these issues may not be important as digestion by proteases can be avoided if the delivery is accomplished by injection, if the treatment is topical, or if bacteriocins are applied to surfaces (for example to prevent colonisation of surgical implants and catheters).

1.10 Other applications of bacteriocin biology

Aside from usage of colicins as food preservatives and antimicrobial toxins for treatment and prevention of infection, colicins have shown potential to be used as anti cancer drugs[138]. They could also find usage in agriculture to protect crops from pathogens[139,140]. Aside from practical commercial usage, colicins can be great assets to basic microbiology research. Particularly, Rassam and colleagues have used translocation deficient colicin E9 and colicin Ia mutants to label outer membrane proteins BtuB and CirA for fluorescence microscopy[141]. Studying colicin import could provide some insight into the systems which are used by Gram-negative bacteria to provide energy for transport of molecules across the OM. Because of their substantial size, modularity and apparent similarity to natural ligands in terms of binding to specific OM transporter proteins, there is a possibility of creating stalled complexes at various time points during translocation, which can then be studied in detail to elucidate the mechanisms involved in activity of TolABQR and TonB-ExbBD complexes during PMF dependent ligand import.

Aims of this work

Due to their ability to cross the OM, colicins are attractive candidates as species specific delivery vehicles for peptides, compounds or perhaps even proteins which would otherwise be unable to do so. As a proof of concept, the first part of this study aims to deliver a known linear cytotoxic peptide which binds BamB by including it into the disordered N-terminal translocation domain of colicin N. For this and other practical applications, understanding the process of colicin translocation across the OM is very important. Hence the second part of this project focuses on improving our understanding of the colicin reception and translocation processes.

Despite the fact that many structural, biochemical and biophysical studies have been performed on binding of colicins to their native receptor and translocator proteins, with most studies concentrating on biology of colicin E9 as a model colicin, the exact mechanism of translocon formation and subsequent translocation remains unknown. Due to apparent high degree of flexibility and dynamic multi-component nature of the complex as well as limited knowledge of the topology of the outer membrane of *E. coli*, previously used methods cannot provide us with enough information to answer this question. The development of novel chemical cross-linkers and more sensitive and accurate mass spectrometry over recent years enables us to study the topology of *in vitro* and possibly *in vivo* translocon complexes. The second part of this work focuses on the development of a method for in vitro cross-linking mass spectrometry study of colicin translocon complexes as well as interpretation of the data obtained for the model colicin E9 in complex with BtuB, OmpF and TolB. The strategy and most of the methods described in this study can be applied to other, less studied colicins and bacteriocins from different species to understand their translocon formation.

Objectives

- Part 1: Colicin N as a chassis for toxic BamD-binding peptide
  - Design and clone full-length and truncated colicin N constructs
  - Express and purify all constructs
  - Compare cytotoxic activity of these constructs to WT colicin N and its truncated construct without the cytotoxic domain
- Part 2: Studying the topology of the colicin E9-OmpF-BtuB-TolB complex
  - Develop the protocol for the lysine-specific cross-linking experiment
  - Test the existing cross-link identification LC-MS/MS method
  - Identify cross-links in the purified colicin E9(TR) translocon
  - Identify cross-links which occur within the colicin E9 and between colicin E9 and BtuB outside the translocon
  - Validate cross-links according to existing structures and knowledge
  - Use protein docking software to model protein-protein interactions according to cross-links as distance restraints
  - Propose changes to the colicin E9 translocon model according to findings.

2. Materials and Methods

2.1 Strains, media and growth conditions

For transformation and plasmid preparation, *E. coli* DH5α cells were grown in lysogeny broth (LB) media with appropriate antibiotic in a shaking incubator overnight at 37°C. Agar plates for colony isolation were also made on the basis of LB media. All colicin N-TR wild type and mutant constructs were expressed in BL21 (DE3) and grown in LB with 100 mg/L of ampicillin, with 500 μM Isopropyl β-D-1-thiogalactopyranoside (IPTG) during induction in a shaking incubator at 37°C, and 160 rpm. Full-length colicin N and its mutant constructs were expressed in terrific broth with addition of 100 mg/L of ampicillin in a shaking incubator at 37°C, 100 rpm. 500 μM IPTG was added during induction.

Colicin growth inhibition assays were conducted on firm LB and M9 agar plates overlayed with inoculated soft LB and M9 agar. The overnight cultures for this experiment were grown in LB, while day cultures (inoculum for soft agar) were grown in the same media as used for the plates. See tables 2.1.1 and 2.1.2 for the list of strains and media recipes.

Table 2.1.1 Bacterial strains

Strain	Genotype	Source
JM83	*ara, Δ(lac-proAB), rpsL,* (Str^R)*[φ80 dlacΔ(lacZ)M15], thi*	Kleanthous lab stock
B^E3000	*$OmpC^-$*	Kleanthous lab stock
BL21(DE3)	*fhuA2, lon^-, $ompT^-$ gal (λ DE3), dcm, ΔhsdS*	Novagen
T7 Express *lysY/I^q*	*fhuA2, lon^- $ompT^-$ gal, lysY, $lacI^q$, $McrA^-$, $McrBC^-$, $EcoBr^-$ m^-, Mrr^-*	NEB
DH5α (NEB5α)	*fhuA2, Δ(argF-lacZ)U169, phoA, glnV44, Φ80Δ (lacZ)M15, gyrA96, recA1, relA1, endA1, thi-1, hsdR17*	NEB

Table 2.1.2 Media recipes

	Components	Additives (final concentration)	Sterilisation method
LB	For 1L: 10 g Peptone 5 g Yeast Extract 5 g NaCl	Ampicillin (1X) – if required IPTG(1X) – if required	Autoclaved (Media cycle)
M9	For 1L or 100ml 10X: 12.8 g Na_2HPO_4-$7H_2O$ 3 g KH_2PO_4 0.5 g NaCl 1.0 g NH4Cl	0.4 % glucose 2mM $MgSO_4$ 0.1 mM $CaCl_2$ 0.2% Casamino acids	Salts and additives autoclaved separately and mixed before use
LB-Agar	For 1L: 1L LB broth 12g Agar powder	Ampicillin (1X) – if required	Autoclaved (Media cycle)
M9-Agar	For 1L: 1L dH_2O 12g Agar powder	M9 salts (1X) 0.4 % glucose 2mM $MgSO_4$ 0.1 mM $CaCl_2$ 0.2% Casamino acids	Water with agar was autoclaved separately and mixed with salts and additives before use
SOC Media	For 1L: 20 g Tryptone 5 g Yeast Extract 0.6 g NaCl 0.2 g KCl	0.4 % Glucose	Autoclaved (Media cycle)
Ampicillin (1000X)	For 10 mL: 1g Ampicillin powder	-	Filter sterilised (0.22 μm)
IPTG – 0.5M (1000X)	For 10 mL: 1.2g IPTG powder	-	Filter sterilised (0.22 μm)

2.2 Generation of mutant colicin N plasmids

The list of plasmids used in this work is shown in table 2.2

Table 2.2 Plasmids used in this work

Plasmid	Expression Strain	Resistance	Source
pET21a-ColN-WT	BL21(DE3)-LysY	Ampicillin	pColN-WT Kleanthous lab
pET21a-ColN-TR (Δ186-387)	BL21(DE3)	Ampicillin	pKBJ11 Kleanthous lab
pET21a-ColN-^{33}SAGIALQW^{40}	BL21(DE3)-LysY	Ampicillin	This work
pET21a-ColN-^{80}SAGIALQW^{87}	BL21(DE3)-LysY	Ampicillin	This work
pET21a-ColN-TR-^{33}SAGIALQW^{40}	BL21(DE3)	Ampicillin	This work
pET21a-ColN-TR-^{80}SAGIALQW^{87}	BL21(DE3)	Ampicillin	This work

2.2.1 Primer design

Two pairs of mutagenic primers for PCR, one per each insertion position, were designed using SnapGene software to create identical insertion mutations in both parent plasmids (pColN-WT, and pKBJ11). Forward and reverse primers of each pair were designed to anneal back-to-back either side of the peptide insertion position and have a melting temperature within 5° C of each other. In addition to the 18-20 nucleotides of annealing sequence, forward and reverse primers each included 12-nucleotide overhang corresponding to frequently used *E. coli* codons for 4 amino-acid-long halves of the SAGIALQW sequence at 5′ ends. The melting temperature was calculated for the annealing portion of the primers without the overhang with the New England Biolabs (NEB) Q5 temperature calculator (http://tmcalculator.neb.com/). This strategy is frequently used with and recommended by NEB Q5 site-directed mutagenesis kit to generate small to medium insertion mutations. Primers were synthesised by eurofins genomics and shipped dry. Upon arrival they were resuspended in nuclease free water at 100 pM/µl prior to first

use and stored at -20° C thereafter. For primer sequences and annealing temperatures see table 2.2.1.

Table 2.2.1. Primer sequences and melting temperature

Name	Sequence	T_m
^{33}SAGIALQW^{40} Forward	5′-GATACCTGCAGAACCATTACTTCCTGCGCC-3′	65° C
^{33}SAGIALQW^{40} Reverse	5′-GCCCTGCAGTGGAGTGCATCAAGTAACCGAGG-3′	64° C
^{80}SAGIALQW^{87} Forward	5′-GCCCTGCAGTGGAGTGGAAATAGAGGTAATAATGGTGATG-3′	64° C
^{80}SAGIALQW^{87} Reverse	5′-GATACCTGCAGAATTCCCACCAGGTTTAGGC-3′	65° C

2.2.2 Polymerase chain reaction (PCR)

All PCR reaction mixtures contained 20 ng of template DNA (1 µl), 1.25 µl of forward and reverse primers at 10µM, 12.5 µl of Master Mix (Q5 Site-directed mutagenesis, NEB), 9µl nuclease-free water. The reaction was carried out in Eppendorf Master cycler as follows:

initial denaturing step at 98° C – 2 min

25 cycles { denaturing step at 98° C – 30 sec; annealing step at 65° C – 30 sec; extension step at 72° C – 4 min

final extension at 72° C – 10 min

After amplification, the reaction products were treated with kinase, ligase and endonuclease mixture (KLD) to circularise the product and degrade the parent DNA to minimise the chance of transforming *E. coli* with the parent plasmid. The reaction was set up as follows: 1 µl of PCR product, 5µl of KLD buffer (NEB), and 1µl of KLD enzyme mix (NEB), and 3 µl of nuclease-free water. The mixture was incubated at 37° C for 1 hour, after which, 5 µl were used to transform DH5α (NEB) competent cells. DNA of transformants purified according to the protocol described above and checked for presence of the desired mutation

via sequencing using standard T7 promoter and terminator primers and DNA at concentration 100ng/µl.

2.2.3 Competent cell preparation

After plating colonies from an aliquot of NEB5α (or BL21 or LysY) cells and incubating them over night, one colony was picked to start a 5 ml over night culture in LB. This culture was then used as an inoculum for a 50 ml culture in LB, which was grown to OD_{600} of 0.6. Cells from this culture were first pelleted by centrifuging at 3,000 x g at 4° C for 10 minutes. The cell pellet was then resuspended in 20 ml of ice-cold sterile 20 mM Tris pH 8.0 with 50mM $CaCl_2$ and incubated on ice for 1 hour. The suspension was then centrifuged at 4° C using the same settings and rotor as before, the supernatant discarded and cell pellet resuspended in 2 ml of sterile 20 mM Tris pH 8.0 50mM $CaCl_2$ with 20% Glycerol at 4° C. The resulting suspension was divided into 100 µl aliquots, which were subsequently flash frozen by immersion into liquid nitrogen and stored at -80° C until use.

2.2.4 Transformation of chemically competent cells

2µl of plasmid DNA purified using QIAGEN miniPrep kit or 5 µl of a PCR/KLD reaction product were used in all transformation procedures. After addition of the DNA, cells were incubated on ice for 1 hour and subsequently heat-shocked by immersing the tubes into a 42° C water bath or a heat block at the same temperature for 45 seconds or 10 seconds for BL21 cells. After the heat shock cells were allowed to recover on ice for 15 minutes before addition of SOC media and incubation at 37° C for 45 minutes. After the final incubation 50 µl of cells (200 µl if the PCR mixture were used for the transformation) were plated onto the petri dishes with LB agar with selection agent (100 µg/ml Ampicillin for pET21a

plasmids). Plates were allowed to dry before placing them upside down into the 37° C incubator over night.

2.2.5 Purification of plasmid DNA

5 ml culture of transformed DH5α cells were grown in LB supplemented with 100 μg/ml Ampicillin overnight at 37° C. 3 ml of these cultures were pelleted for 5 minutes at 6,000 x g in a table-top centrifuge. The rest of the purification procedure was performed using a QIAGEN mini-prep kit, according to instructions provided by the manufacturer.

2.3 Protein expression and purification

For expression of truncated colicin N constructs (Δ186-387) BL21 cells were transformed with appropriate plasmids and grown in 5 L of LB with 100 μg/ml of ampicillin until OD_{600} reached 0.6-0.7 before adding inducing agent IPTG at 1 mM final concentration. For expression of wild-type colicin N, lysY cells (NEB) were transformed with appropriate plasmids and grown in 5 L of terrific broth (TB) with 100 μg/ml ampicillin until OD_{600} reached 1.2 before addition of IPTG at 1 mM final concentration. After 2 hours of incubation at 37° C, cultures were pelleted for 25 minutes at 3,000 x g. Cell pellets were stored at -20° C. To purify the required proteins, a combination of immobilised metal affinity chromatography (Ni-IMAC) and size exclusion chromatography (SEC) were used.

2.3.1 Ni-IMAC

Immediately before purification, cell pellets were thawed on ice, resuspended in 25 ml Ni-IMAC wash buffer containing 50 mM Tris pH 7.5, 300 mM NaCl, 5mM imidazole, and 300 μl of protease inhibitor phenylmethane sulfonyl fluoride (PMSF). The suspensions

were then sonicated on ice in 10 second bursts at 30 second intervals for 1.5 minutes of total sonication time. This was repeated once more after the cells were cooled on ice for 10 minutes. The soluble cell lysate was then separated from cell debris by centrifuging for 30 minutes at 15,000 x g and filtered through 0.45 μm syringe filter before being loaded onto the Ni His-Trap HP (GE Healthcare) column using ÄKTA Prime purifier at 4° C and flow rate of 1 mL/min. The rest of the IMAC purification procedure was also performed on ÄKTA Prime as follows. The column was washed with 20 x CV (column volumes) of wash buffer described above, and the protein eluted with a gradient of imidazole from 5 to 252.5 mM over 100 ml, while monitoring absorption at 280 nm. Peak fractions were collected and analysed on an SDS-PAGE to confirm the identity of the purified protein and assess its purity in the eluate. Fractions with the highest protein of interest to contaminant ratio were pooled and dialysed into SEC buffer overnight (4 L 25mM Tris pH 7.5, 150 mM NaCl).

2.3.2 Size exclusion chromatography

After Ni-IMAC and dialysis, proteins were concentrated to < 10 ml using Vivaspin 20 concentrators with a 10,000 Da molecular weight cut-off before being loaded onto the size exclusion column.

Superdex S75 26/60 size exclusion column (GE Healthcare) was equilibrated with the SEC buffer described above, then 5-10 ml of concentrated protein was loaded onto the column. It was then washed with the same buffer while monitoring the absorbance at 280 nm. The peak fractions were collected in 4 ml fractions and visualised on SDS-PAGE. Cleanest fractions were pooled before the concentration of each protein was measured as described in section 2.3.4.

2.3.3 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

For separation on the SDS-PAGE, protein samples were added to the 4x SDS-PAGE loading dye: 200mM Tris-HCl, pH 6.8, 8% (w/v) SDS (electrophoresis grade), 0.4% (w/v) bromophenol blue, 40% (w/v) glycerol, 400mM β-mercaptoethanol. Samples were boiled at 98° C for 5 minutes prior to loading onto the gel alongside the unstained 14.4 to 116 kDa molecular weight markers (Fermentas #SM0431). Unless otherwise stated, the stacking gels contained 5% bis-acrylamide, and the resolving gels contained 12% bis-acrylamide. The gels were run in the Hoefer "Mighty Small" tank at constant current of 30 mA per gel for 45 minutes in Tris-glycine running buffer: 0.25M Tris, 1% SDS, 1.92M glycine. Proteins were visualised by staining the gel with 0.2% (w/v) Coomassie brilliant blue R250 (Pierce) in 10% v/v acetic acid and 10% v/v ethanol until protein bands were visible and de-stained in the same solution but without the dye.

2.3.4 Protein concentration determination

The concentration of all proteins used in this study was calculated according to the Beer-Lambert law:

$$A = \varepsilon c l \Longrightarrow c = \frac{A}{\varepsilon l}$$

Where A is the absorbance of the solution at 280 nm, ε is the molar extinction coefficient (M^{-1} cm^{-1}), c is the concentration of proteins (M^{-1}), and l is light path (cm).

The absorbance of a at 280nm was measured using Eppendorf Bio Photometer, the length of the light path is 1 cm, and the extinction coefficients were calculated using the ProtParam tool (see table 2.3.4)

Table 2 Extinction coefficients of proteins and complexes used in this study

Protein	Molecular weight	Extinction coefficient ε
Colicin N-WT	42807.13 Da	50880 M^{-1} cm^{-1}
Colicin N-TR (Δ186-387)	20914.73 Da	26930 M^{-1} cm^{-1}
Colicin N-WT ^{33}SAGIALQW^{40}	43652.1 Da	56380 M^{-1} cm^{-1}
Colicin N-WT ^{80}SAGIALQW^{87}	43652.1 Da	56380 M^{-1} cm^{-1}
Colicin N-TR ^{33}SAGIALQW^{40}	21759.7 Da	32430 M^{-1} cm^{-1}
Colicin N-TR ^{80}SAGIALQW^{87}	21759.7 Da	32430 M^{-1} cm^{-1}
BtuB-Colicin E9-OmpF-TolB translocon	270934.73 Da	398720 M^{-1} cm^{-1}

2.4 Colicin activity test

Soft LB or M9 agar kept at 45° C was inoculated with approximately 10^8 cells of mid-log phase *E. coli* B^E3000 or JM83 per 10 ml and poured onto standard high salt LB or M9 agar plates. After the plates set, 2.5 µl drops of serial dilutions of wild type colicin and its mutant varieties spanning a concentration range of 10 µM down to ~ 50 pM were spotted onto each plate. Spots were allowed to dry before plates were placed in the incubator at 37° C overnight. Colicin N-WT and colicin N-TR (Δ186-387) proteins used as controls for this assay were generously provided by K. Jansen.

2.5 Lysine-specific cross-linking of ColE9 translocon

The translocon was purified by N. Housden (Kleanthous lab), and contains colicin E9, OmpF, BtuB, and TolB proteins with the following modifications: Colicin E9 is present as a C-terminally His-tagged truncated protein lacking the nuclease domain with 3 cysteine residues at positions 33, 85 and 104, while TolB has a P201C mutation. These mutations

stabilise the translocon so it can survive purification by allowing formation of disulphide bonds between residues 85 and 104, which prevents unfolding of the of ColE9 translocation domain, and between residues 201 of TolB and 33 of ColE9, which tethers TolB to ColE9[33,74]. For cross-linking experiment, the translocon was dialised into PBS buffer at pH 7.4 with 50 mM β-Octyl glucoside. Final concentration of the complex was measured at 8.5 μM using the method described in section 2.3.4. The complex was then stored at 4° C.

2.5.1 Cross-linking reaction

BS2 and BS3 crosslinkers (Thermo Fisher) were dissolved in water at 25 mM and added to protein to a final concentration of 0.5 mM immediately prior to the experiment. The first test reaction was allowed to proceed for 30 minutes, however in subsequent experiments this was shortened to 5 minutes, before quenching the reaction with 50-100 mM Tris NaCl Buffer at pH 7.4.

2.5.2 2-D Blue-Native PAGE

To avoid identifying the crosslinks between different translocon complexes, the cross-linked products were first separated on 4-16% Blue Native (BN) PAGE (Life Technologies) for 2 hours at 4° C and constant voltage 150 V. After the first separation, the bands with molecular weight similar to that of the non-cross-linked complex were cut out, boiled in SDS-loading buffer and loaded into the wells of the 10% SDS-PAGE (1.5 mm thick) and run as a regular SDS-PAGE (see section 2.3.3) with increased current (60 mA).

2.5.3 In-gel trypsin digestion

Appropriate bands were excised, destained with 50% acetonitrile (ACN) and placed in 50 mM triethylammonium bicarbonate (TEAB) for in-gel tryptic digest. The proteins were

then reduced with 50 mM TCEP for 30 min, alkylated with 50mM chloroacetamide in 100 mM TEAB for 30 min in the dark, and finally digested with 0.2 μg trypsin (HPLC, proteomics grade) overnight at 37° C. The cross-linked peptides were eluted from the gel with 80% ACN. To prepare samples for loading onto the reverse-phase chromatography column ACN was removed with SpeedVac. Dried samples were resuspended in 5% formic acid (FA).

2.6 LC-MS/MS

Mass spectrometry was performed on Nano-LC coupled to Q Exactive Hybrid Quadrupole-Orbitrap MS (Thermo Fisher). 2.0 μl of the sample was loaded onto an on-line reverse-phase chromatography column at 10 μl/min. Subsequently, the peptides were eluted with a 1-hour solvent gradient (0-30% ACN) at 200 nl/min and subjected to electrospray ionisation in positive ion mode. The parent ions with m/z in the range 350-2000 were analysed in the quadrupole, after which the top-10 parent ions in each fraction with a charge state $\geq 3^+$ were fed into the high-energy collisional dissociation (HCD) chamber for fragmentation. The fragment ions were subsequently passed through the C-trap to the orbitrap for analysis (first m/z = 100).

2.7 Mass spectrometry data analysis

The data obtained from MS/MS runs was analysed with MaxQuant[142] to determine the protein composition of the samples by searching the spectra against the full *E. coli* proteome database with updated sequences of the proteins involved in formation of the translocon and common contaminants. The mass tolerance for the precursor ions during the first search was set at 50 ppm, acetylation of the N-terminus and oxidation of methionine were selected as variable modifications. The results of this search were used to make up the

small database for pLink[143] searches for cross-linked peptides. To ensure that no proteins of low-abundance in the sample were cross-linked to parts of the translocon during cross-linking reaction, a full database search was performed with pLink as well. In all pLink searches the mass tolerance for the precursor ions during the first search was set at 5 Da, while the mass tolerance for the fragment ions was set at 20 ppm. The results of the search were viewed at 5% false discovery rate (FDR), with an imposed E-value threshold for identified peptides at $5x10^{-4}$, unless otherwise stated, prior to validation.

2.8 Protein-protein docking and model analysis

Identified cross-links were visualized by XLink Analyzer[144] plugin within Chimera[145] and used as unambiguous distance restraints (12±5 Å from Cα of lysines) to model the interaction surfaces between colicin E9 and OmpF with HADDOCK 2.2[146]. As only one cross-linking site was found between OmpF and BtuB, additional restraints were used to model the interaction between the two proteins. These were derived from MD simulations of OmpF-BtuB interactions in lipid environments published in 2015 by Rassam et al.[141].

3. Colicin N as a chassis for toxic BamD-binding peptide

3.1 Introduction

3.1.1 Outer membrane of Gram-negative bacteria

One of the main morphological features that distinguish Gram-negative bacteria from Gram-positive is the presence of an outer membrane: an asymmetrical lipid bilayer with integral proteins. The lipid bilayer of the outer membrane is different to the symmetrical phospholipid bilayer of the inner membrane, which is home to inner-membrane associated proteins and lipoproteins as well as integral membrane proteins. The inner leaflet of the asymmetrical lipid bilayer of the outer membrane is composed of regular phospholipids, the same as those found in the inner membrane. It serves as an anchor to many outer-membrane-associated lipoproteins. The outer leaflet on the other hand is composed of lipopolysaccharide (LPS), which in turn consists of lipid A, core oligosaccharide and O-antigen, and contributes to structural integrity of the cell. LPS is impermeable to certain ions and small molecules, which are vital for bacteria, therefore many integral outer-membrane proteins (OMPs) are embedded into the asymmetric bilayer to act as ion channel and small molecule transporters. Other OMPs include efflux pumps, virulence and adhesion factors, secretion systems, and structurally important proteins. Unlike integral inner-membrane proteins, which are predominantly α-helical and are inserted into the membrane during synthesis, the majority of outer membrane proteins have β-barrel structures with or without N-terminal plug domains and are folded and inserted into the outer membrane only after export into the periplasm[147]. Although some *in vitro* studies show that β-barrel proteins can be spontaneously inserted into lipid bilayers, because it is a very slow process, *in vivo* they usually require a β-barrel assembly machinery (BAM) to catalyse their insertion into the outer membrane[148,149].

3.1.2 The BAM complex

After synthesis, OMP polypeptide chains enter the periplasm through the Sec pathway in an ATP- and PMF-dependent manner, where they interact with periplasmic chaperones and components of BAM, which catalyses their insertion into the outer membrane[147]. BAM-mediated insertion is independent of ATP[150], and is thought to be driven by the free energy of OMP folding in the OM[151].

The BAM complex consists of a 16-stranded β-barrel protein BamA, and four outer membrane-associated lipoproteins BamB, BamC, BamD, and Bam E, although only BamA and BamD are essential for OMP biogenesis[152,153]. Apart from the C-terminal β-barrel, BamA protein also has 5 N-terminal polypeptide transport associated (POTRA) domains, which are known to interact with substrate polypeptides and other components of the BAM machinery[154,155]. Although the complex has been extensively studied since early 2000s, its first full crystal structure was only published in 2016 by Han and colleagues (see figure 3.1.2)[156].

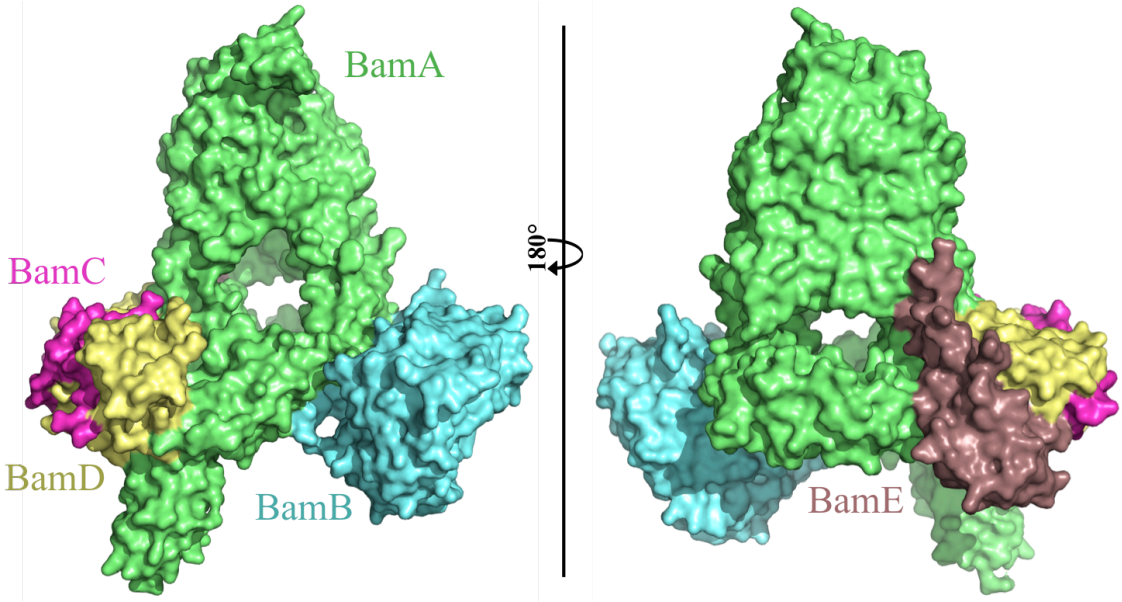


Figure 3.1.2 Crystal structure of BAM complex (PDB ID: 5AYW) forming the "top hat" structure, with BamA β-barrel at the "top" part of the structure and accessory lipoproteins BamBCDE bound to POTRA domains of the BamA forming the "hat brim".

While BamA is the largest protein and the catalytic component of the BAM, accessory lipoproteins BamBCDE are involved in regulation of OMP biogenesis[156–159]. BamB is thought to play a role directing the substrate polypeptides to BamA, while BamD is important in substrate recognition process, and BamE is needed for structural stability[154]. The role of BamC is unclear at present, though it is thought to play a regulatory role in substrate recognition by BamD[160].

3.1.3 BamA C-terminal peptide binds to BamD and inhibits growth of *E. coli*

In 2015, Hagan et al. discovered that overexpression of the ^{769}SAGIALQW^{776} fragment of BamA as part of a larger 96 amino acid polypeptide is lethal to *E. coli*[161]. The reason for this toxicity is currently unknown, but it is thought to be due to its ability to bind BamD and as a result competitively inhibit BAM assembly[161]. The SAGIALQW sequence is a variation of a loosely defined motif present in all BAM substrates at the C-terminal β-strand[162–166]. It allows substrate to be recognised by BamD for further insertion into OM. Moreover, some variations of this signal can also be recognised by mitochondrial Sam35 and Sam50 (functional homologue of BamD and BamA) facilitating their assembly at the outer membrane of mitochondria by Sam50[167–170].

3.1.4 Experimental design

Using the original 90 amino acid peptide containing SAGIALQW, which comes from the three C-terminal β-strands of BamA would be problematic due to the fact that it would be too long to chemically synthesise, but also challenging to express in sufficient quantity in *E. coli* due to its toxic effect. Moreover, its propensity to form β-sheets would likely prevent it from entering cells. An alternative approach could be to chemically synthesise the active 8-amino-acid peptide, which would also be able to cross the membranes through

porins. However, this option is also not without issues: the short peptide may be subject to bacterial natural defenses, and due to poor sequence selectivity of BamD and Sam35 it may be capable of inhibiting the mitochondrial SAM machinery as well, which is undesirable if the peptide is to be used as a treatment for infection.

Since colicin N has a long intrinsically disordered sequence at its N-terminus of approximately 90 residues long, which is capable of selectively crossing the outer membrane of *E. coli*, it seemed like an ideal candidate for use as chassis to transport the SAGIALQW peptide into the cell. To make sure the transport across the membrane is most efficient, and to prevent the export of the peptide by efflux pumps and other cellular machinery, it made sense to use the entire protein. However, because the full-length colicin N is an efficient pore-forming toxin, its cytotoxic domain was removed to allow the assessment of cytotoxic activity of SAGIALQW peptide. This way, the peptide was inserted into the N-terminus of the colicin N-TR construct (ColN Δ186-387). This section of colicin N has however been reported as toxic to *E. coli* B^E3000 cells[56].

The disordered region of the translocation domain of colicin N has important functional elements: at least one OmpF binding site (OBS) and a TolA-box sequence. As these are essential for colicin N translocation, to avoid any translocation issues two positions were chosen for SAGIALQW insertion: one between residues 32-33 with OBS and TolA-box sequence flanking the peptide insert, and another half way from the last TolA-box residue to the beginning of ordered structure between residues 79-80 (see figure 3.1.4). The ability of resulting colicin N-TR constructs to translocate is crucial for SAGIALQW to have any effect on the cells. To test this ability, full-length colicin N constructs with inserts in the N-terminus were made too. If the insertions had any effect on translocation efficiency we would expect to see a decrease in mutant colicin N activity compared to wild-type.

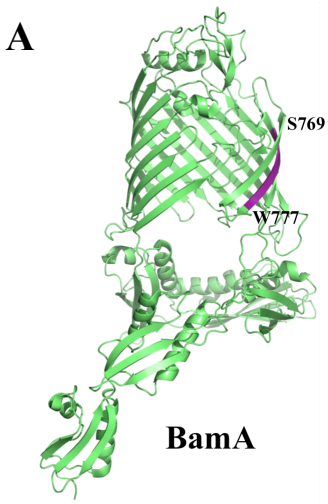


Figure 3.1.4 A. Crystal structure of BamA including all POTRA domains (PDB ID: 5AYW). SAGIALQW sequence on the 14th β-strand of the β-barrel is coloured purple. **B.** Sequence of disordered N-terminal domain of colicin N. OmpF binding site and the TolA-box are highlighted green and yellow respectively. Sites of the SAGIALQW insertion are shown with purple arrows.

B

^{1}MGSNGADNAH NNAFGGGKNP GIGNTSGAGS NGSASSNRGN SNGWS

WSNKPHKNDG FHSDGSYHIT FHGDNNSKPK PGGNSGNRGN NGDGS^{90}

3.2 Results

3.2.1 Test expression of Colicin N-SAGIALQW constructs

To test the effect of SAGIALQW insertion into the disordered N-terminal region the constructs were made by mutagenesis PCR using the full-length colicin N and its C-terminally truncated variant colicin N Δ186-387 in pET21a plasmids as templates using the methods described in section 2.2.

Since all truncated constructs were thought to be harmless to the expressing cells, they were expressed using the standard *E. coli* T7 expression strain BL21 (DE3). Constructs based on the full-length colicin N would contain a pore-forming domain and therefore would likely be toxic to the expressing strains and would inhibit the growth of the entire culture if released into the medium through spontaneous lysis. Therefore, T7 Express LysY/I^{q} cells were used instead of BL21 (DE3), as they allow for tight control of gene expression through *lacI*^{q}, and are less likely to lyse due to having LacY lysozyme deficient in amidase activity. To test whether changing the strain is enough to produce a reasonable quantity of

proteins, a test expression was done in the same conditions as a regular expression, albeit in smaller volume. The results are shown in figure 3.2.1 A.

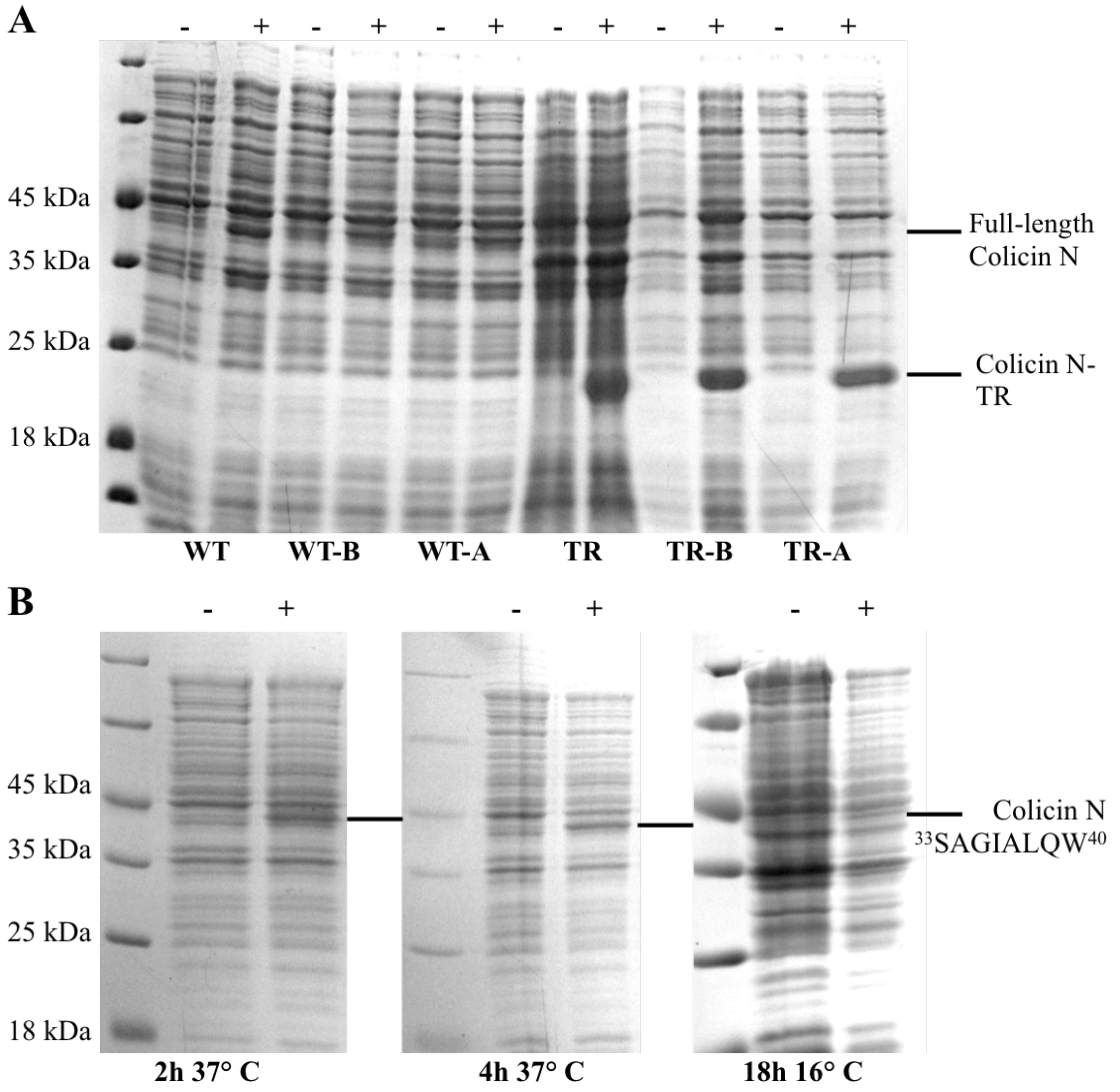


Figure 3.2.1 A SDS-PAGE of the test expression of all mutant constructs as well as colicin N-WT and colicin N-TR (Δ186-387) in LB. The mutants with SAGIALQW insertion at residues 33-40 are labelled 'WT-B' and 'TR-B', and those with the insertion at residues 80-87 are labelled 'WT-A' and 'TR-A' depending on the length of the original construct. Samples before and after induction are labelled '-' and '+' respectively. **B** Test expression of full-length colicin N with insertion at residues 33-40 in terrific broth for 2, 4, and 18 hours. The samples before and after induction are labelled same as above.

While the truncated variants of the colicin N were expressed well, the full-length constructs could not be seen on the stained SDS-PAGE gel, which led to a change of strategy used to express these constructs. The growth media was changed to terrific broth to allow higher

cell density before the induction to increase the overall protein yield. Three growth conditions were tested: 2h at 37° C, 4h at 37° C, and 18h at 16° C. The results are shown in figure 3.2.1 B. Since the 4h induction did not provide a significant increase in protein yield over the 2h induction and the overnight expression at 16° C failed to show a significant increase in colicin N compared to non-induced cells, the large-scale expressions were carried out for 2h at 37° C.

3.2.2 Full-scale expression and purification of colicin N-SAGIALQW constructs

All constructs were expressed in 5L cultures. The cells were harvested, lysed, and the proteins purified using a combination of Ni-affinity and size exclusion chromatography to achieve a satisfactory degree of purity as described in section 2.3. Stained SDS-PAGE gels of peak fractions after each round of purification of each protein are shown in figure 3.2.2 indicating the pooled fractions which were used for further experiments.

3.2.3 Effect of the SAGIALQW peptide insertion on activity of full-length colicin N

To investigate the effect of insertion of the 8-amino acid peptide on translocation of the protein, the cytotoxic activity of full-length constructs were compared with that the wild-type colicin N. JM83, a derivative of *E. coli* K-12 strain susceptible to colicin N, was used in the initial colicin activity test described in the materials and methods section 2.4. The minimal inhibitory concentration (MIC) for each protein was determined by the minimum concentration of the protein found to generate a zone of clearing in the soft agar inoculated with sensitive cells. All assays were done in triplicate, and representative plates are shown in figure 3.2.3. The MIC of colicin N-WT determined from these experiments was 169 pM, colicin N ^{33}SAGIALQW^{40} – 14 nM, colicin N ^{80}SAGIALQW^{87} – 57 pM. The result for the wild-type colicin N is similar, albeit slightly lower than the previously reported MIC (~500

pM-1 nM)[171]. This could be due to use of a different strain of *E. coli* JM83, a K-12 derivative used in this study versus the sensitive K17 used by Sharma et al.[171].

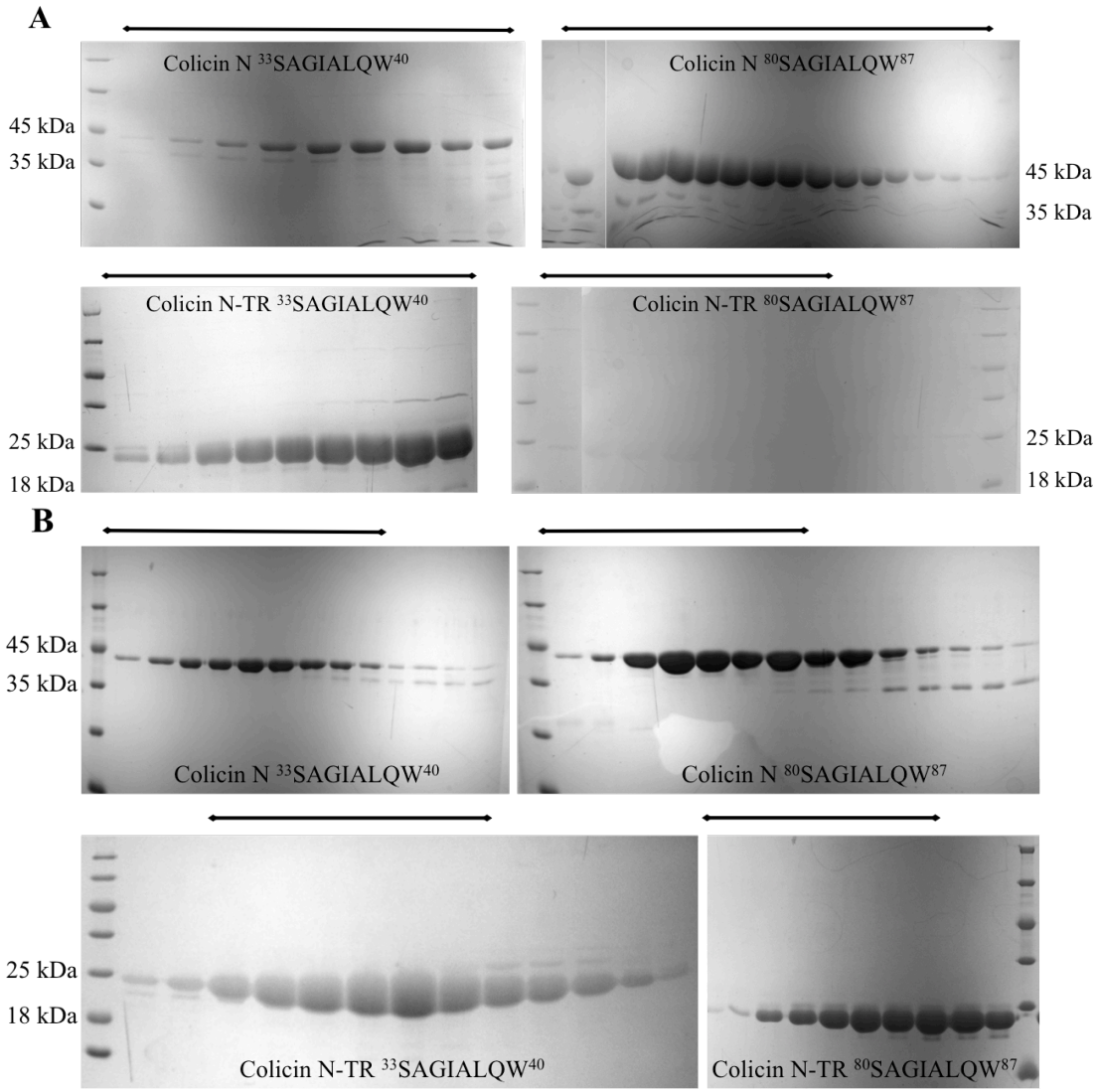


Figure 3.2.2 A. SDS-PAGE of Ni-IMAC elution fractions of all mutant constructs. Pooled fractions for size exclusion chromatography are marked with arrows. **B.** Elution fractions from the gel filtration size exclusion chromatography column, pooled fractions for colicin activity test are marked with arrows.

It is apparent from the MIC of full-length colicin N with the SAGIALQW insert at residues 33-40, had increased approximately 100-fold, while the MIC of the construct with the insert at residues 80-87 is in the similar range to the wild-type, even reduced by up to 3 fold.

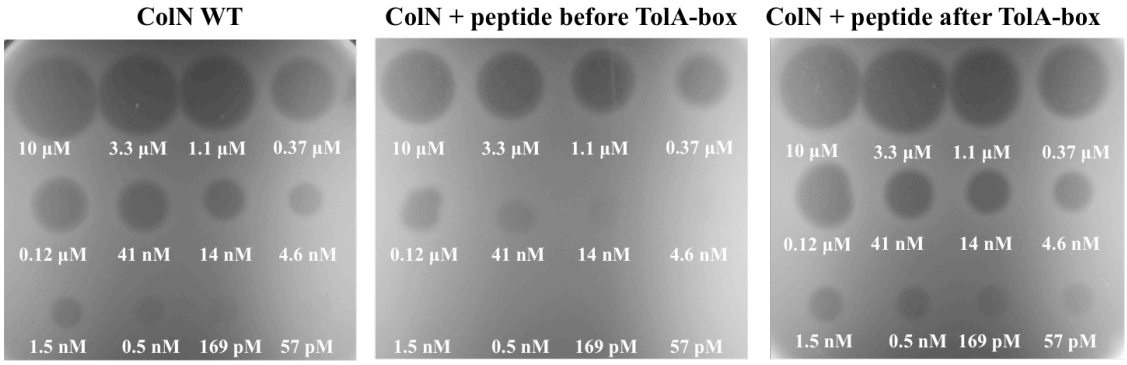


Figure 3.2.3 Growth inhibition assay using full-length colicin N constructs with and without the BamD binding insert on LB plates using sensitive JM83 cells. The concentration of proteins spotted onto the agar plates is indicated below the spots.

Since the insertions are in the intrinsically disordered N-terminus, they are unlikely to have any effect on the overall protein folding and the ability of the C-terminal pore domain to form pores in the inner membrane. The most likely reason for the reduced activity of the ColN-^{33}SAGIALQW^{40} construct is impaired translocation of the protein into the periplasm. This could be due to reduced affinity of the TolA-box to TolA, as it may affect the structural rearrangement of TolA-box residues 40-67 upon contact with $TolA_{III}$, which would now be at positions 48-75. This region was previously shown to have propensity to form β sheets upon binding to $TolA_{III}$[172–174], so it is possible that the inserted peptide (β strand of BamA) may interfere with any structural rearrangement required for this to occur. Another possible reason for the reduced activity of colicin N with the insert at residues 33-40 may be reduced accessibility of the colicin TolA-box to $TolA_{III}$, as the number of residues between the OBS site and the TolA box in this construct is increased by approximately 20% due to addition of the BamD epitope. Consequently, the TolA-box may be too close to the OM or partly obstructed by the porin to easily contact $TolA_{III}$.

The insertion of the peptide after the TolA-box of the colicin N had no effect on colicin translocation and overall activity, which is unsurprising as its TolA-box was located in the same place as that of the wild-type and the insert was unlikely to disrupt its interactions with TolA and OmpF.

To confirm these findings, an assay comparing the growth of JM83 cells in presence of wild-type colicin N and colicin N with the insert at positions 80-87 in liquid culture was performed. However, it could not be completed due to the fact that the initially sensitive bacteria have developed tolerance to colicin N within 1-2 hours after addition of proteins at final concentration 45-90 nM (data not shown) presumably due to a change in the level of OmpF expression at the outer membrane. For this reason, B^E3000 cells, which do not have a gene that encodes a functional homologue of OmpF – OmpC, were used in further experiments involving the colicin N-TR constructs.

3.2.4 Suitability of JM83 and B^E3000 cells for assessment of colicin $N_{\Delta 186\text{-}387}$

Following the observation that JM83 cells may not be suitable for the colicin N activity tests, the difference between the activity of colicin N TR (Δ186-387) construct without the insert was tested using both JM83 and B^E3000 cell lines. None of the tested concentrations of the truncated colicin N had any cytotoxic effect on the JM83 cells, while some were inhibitory to B^E3000 cells as shown in figure 3.2.4. The minimum inhibitory concentration of the colicin N-TR towards the B^E3000 cells was estimated at ~ 1-1.5 μM, which is 10-fold less than previously reported by Johnson et al.[56]. The reason for this discrepancy is not clear. However, due to differences in protein expression protocols used in both labs, it is possible that the degree of Asn4 deamidation in colicin N-TR constructs, which occurs during prolonged expression times (over 2 h) and is thought to inhibit the activity of colicin N-TR, is greater in our protein sample[56].

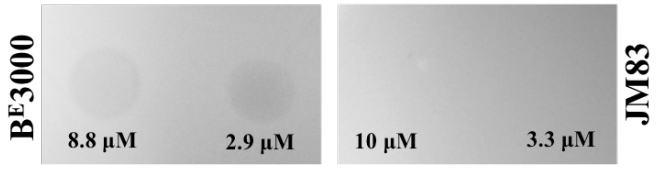


Figure 3.2.4 Colicin N (Δ186-387) growth inhibition assay using soft agar inoculated with either B^E3000 or JM83 cells. The concentration of the colicin N-TR solution, which was spotted onto the plates, is marked below the spots.

3.2.5 Effect of different media on colicin $N_{\Delta 186\text{-}387}$ toxicity in B^E3000 cells

The reason for the observed "fuzziness" of the clearing zones (fig. 3.2.4) exhibited by colicin N-TR construct is unknown. Interestingly, performing the assay on minimal media plates seems to partially alleviate the issue and results in clearer zones even at higher concentrations of the protein as seen in figure 3.2.5. The assay was done in triplicate and representative results are shown.

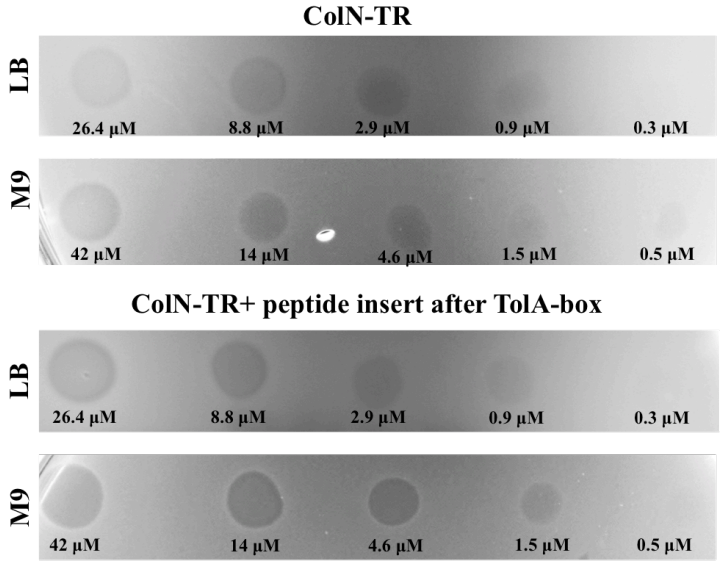


Figure 3.2.5 Colicin N-TR growth inhibition assay comparison between two different media types: rich (LB) and minimal (M9).

The effect of the medium is more pronounced for the constructs with the BamD epitope insert, however the reason for this is not clear at present. It could be that the composition of the outer membrane, including the level of OmpF expression is different when cells are grown in M9 medium compared to the rich LB medium and allows more molecules to bind and translocate using OmpF. However, this hypothesis was not tested in this study and there is limited literature regarding this issue.

3.2.6 Effect of the peptide insertion on the activity of truncated colicin $N_{\Delta 186-387}$ in BE3000 cells

The cytotoxic activity of truncated colicin N constructs with and without the BamD binding inserts was compared by the spot test on M9 agar plates. As with other spot tests, this assay was conducted in triplicate and a representative plate is shown in figure 3.2.6.

The original truncated colicin N showed MIC ~1.5 μM, while the MIC of the truncated colicin N construct with the insert at residues 80-87 was ~0.5-1.5 μM, and the construct with the insert at residues 33-40 failed to show any inhibitory activity. As with the full-length colicin, insertion of the SAGIALQW sequence into the region between OBS and TolA-box, resulted in decrease in activity of the construct. Because the observed effect of this insertion for full-length colicin was a 100-fold increase in MIC, and the upper limit of the protein concentration tested in this study was 50 μM, and not 150 μM, it was impossible to detect any activity of this construct.

The effect of the BamD epitope insertion after the TolA box of colicin N-TR on its activity is unclear. Although the zones of clearing are much more pronounced and defined than those created by the original colicin N-TR, the MIC of the construct with the insertion

remained unchanged. This suggests that unlike the full-length colicin N, which only requires a small number of translocated molecules to kill a single cell, the construct with the BamD epitope insert, like the unmodified colicin N-TR, must require many more to accomplish the same task, and could act to exacerbate the effect caused by the colicin N-TR domain itself. This is consistent with the fact that there are many BAM complexes, which work in concert to support the cell growth, and therefore the inhibition of activity of only a few is unlikely to have a significant effect on cell viability.

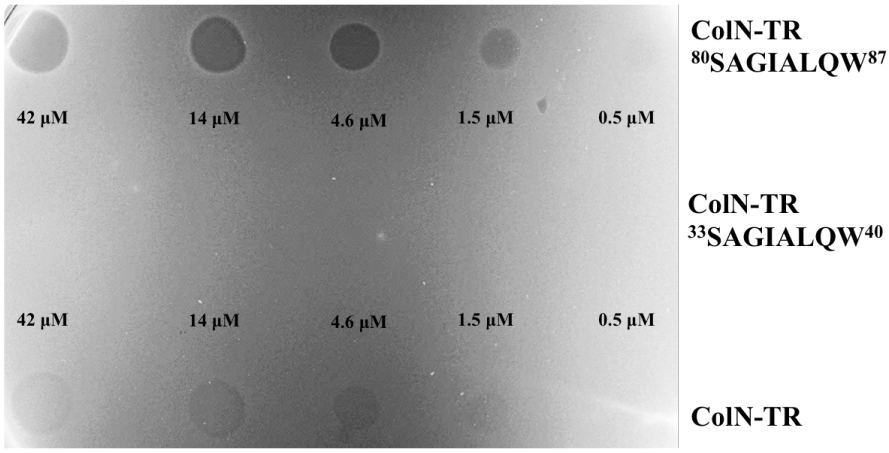


Figure 3.2.6 Colicin growth inhibition assay comparison between the truncated colicin N constructs: ColN-TR (Δ186-387), ColN-TR ^{33}SAGIALQW^{40}, and ColN-TR ^{80}SAGIALQW^{87} on the M9 agar plates. The concentrations of protein solutions used for this assay are marked between the rows of 5 spots.

However, if deamidation of asparagine at position 4 prevents the cytotoxic effect of the colicin N-TR, then many of the translocated colicin molecules would be inactive in this regard, so would have limited capacity to induce lysis, which is evidenced by partial clearings in the colicin growth inhibition assays. While the presence of the BamD binding peptide as part of the disordered N-terminal region of the protein in the periplasm may bring about the cytotoxic effects if enough molecules are accumulated, overcoming the need for intact Asn4.

3.3 Discussion

Colicin E9 N-terminus has been used previously to deliver a TolB-binding peptide to the periplasm[175]. The data collected in this study suggest that colicin N TR domain can be used to deliver other functional epitopes to the periplasm. However, of the two original insertion positions, only the insertion of the peptide at residues 80-87 of the disordered N-terminus of the colicin can be tolerated without changes in efficiency of colicin transport across the membrane. Due to mild cytotoxicity of the colicin N-TR construct, the effect of import of BamD epitope could not be easily elucidated. Although the addition of the BamD epitope to the colicin N-TR construct after the TolA-box showed a mild improvement in activity over the regular truncated colicin N, the reason for this improvement is unclear. Elucidation of the cause of this improvement as well as the investigation of the effect of the minimal media on outer membrane composition would require extensive and complex testing including probing outer membrane protein expression in different media, design of other constructs which include the SAGIALQW sequence in the N-termini of other colicins, removing the R domain of the constructs to reduce endogenous cytotoxicity, testing *in vivo* and *in vitro* binding of the construct to BamD.

3.4 Further research questions on colicin N and *E. coli* cell biology

While this project was not successful in establishing a useful application for colicin N-TR domain, it has highlighted some interesting questions, the answers to which are likely to be useful in future antibacterial research. For example, knowing how the surface proteome of bacteria changes depending on the environment (which proteins are expressed under which conditions). Understanding which of these changes are essential to the life of bacterial cells may help to target the antimicrobial compounds or proteins towards particular surface receptors and effectors depending on the environment of the target bacteria.

4. Probing the interface between colicin E9 and its translocon proteins by crosslinking mass spectrometry

4.1 Introduction

The recognition of colicins at the bacterial surface and the mechanism of their translocation across the outer membrane have been extensively studied for over twenty years. We now know that colicins bind to their primary receptor proteins with high affinity using their central R domain and then probe the nearby area for their translocator protein using their unstructured N-terminal T domain. After binding to their translocator proteins, the disordered N-termini of the colicins cross the outer membrane and contact their second translocator protein, which transfers the energy for unfolding and translocation of the rest of the molecule. Colicins E3, E9 and Ia are some of the most studied colicins in this regard. Crystal structures of each of the receptor and translocator proteins with segments of colicin E9 are known and shown in figure 4.1. The main principles of translocation of these colicins are reasonably well understood but many details remain unclear.

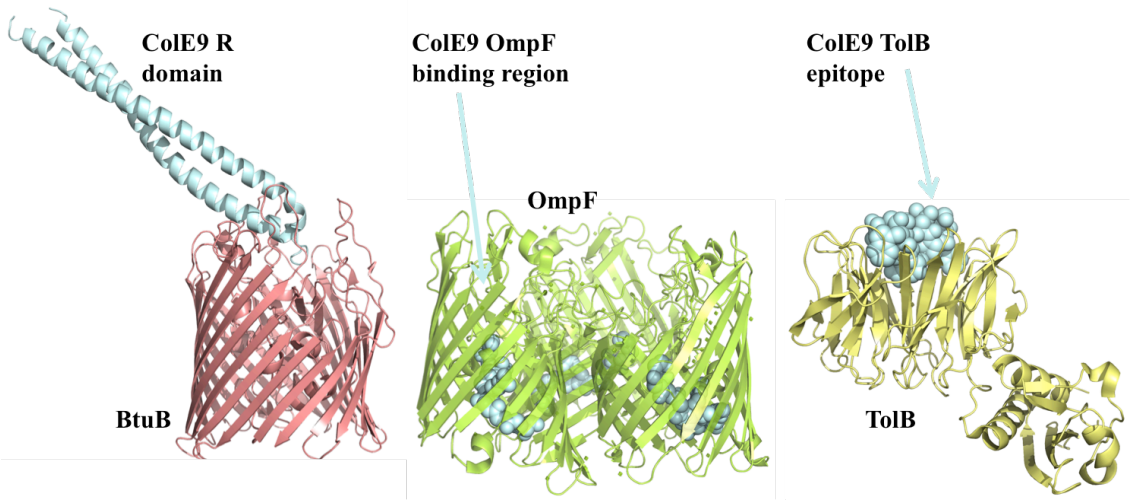


Figure 4.1 Crystal structures of ColE9 translocon components with their binding epitopes. Colicin E9 fragments are shown in cyan as a cartoon model (α-helical hairpin of the R domain) and spacefill (OmpF and TolB binding peptides). PDB ID: 1UJW, 3O0E, and 2IVZ.

The work described in this chapter was focused on applying *in vitro* crosslinking mass spectrometry methods to investigate inter-protein interfaces between the components of the colicin E9 translocon: colicin E9, BtuB, OmpF, TolB. Although the translocon of colicin E9 is one of the most studied and well documented in terms of individual protein-protein interactions and conformation, there are some aspects that are not entirely understood. It is known that colicin E9 binds BtuB prior to binding to OmpF, and that without BtuB colicin E9 cannot translocate. However, it is not known how the colicin quickly locates and recruits available copies of OmpF. It is also known that unfolding of the colicin facilitates release of the immunity protein, but it is not clear what causes this unfolding[74]. The flexibility of the R-domain and movement of the helices of its coiled-coil relative to each other are known to be required for translocation, but it is not clear at what point in the translocation process this occurs[176]. Although crystal structures of colicin E9 with each of the proteins in the translocon have been solved, they provide limited understanding of how the proteins fit together and how the colicin is translocated (fig. 4.1). They also do not provide insight into conformational dynamics, which is of vital importance to the colicins.

Colicins are relatively large proteins, so solving their solution structure and conformational dynamics on a global scale using nuclear magnetic resonance (NMR) or electron spin resonance (ESR) is also problematic. Attempts at crystallization and cryo-electron microscopy of whole translocons so far have been unfruitful. Recent advances in chemical crosslinking of protein complexes as well as advances in mass spectrometry technology have made it possible to study topologies of multi-protein complexes[177–181]. In this study we combine chemical crosslinking and mass spectrometry to study the topology of the translocon complex as well as the events, which are responsible for its formation.

4.1.1 Chemical crosslinking for probing protein-protein interactions

Chemical cross-linking of proteins, along with Förster resonance energy transfer (FRET), yeast two-hybrid system, co-immunoprecipitation, and tandem affinity purification are important tools for studying protein-protein interactions *in vivo* and *in vitro*. The advantage of using chemical cross-linking over the other methods is the sensitivity to transient interactions by stabilising protein complexes, which can then be studied by a variety of biochemical and biophysical methods such as SDS-PAGE, Western blotting and tandem mass spectrometry to identify interaction partners. Many different chemical cross-linking reagents have been developed over the years with varying length and functional groups for a range of applications. They can range from zero-length, homo-bifunctional, non-specific, cleavable crosslinkers (formaldehyde) to long, hetero-bifunctional, specific to amine and sulfhydryl groups, non-cleavable cross-linkers (LC-MSCC). Some crosslinking reagents can also be genetically encoded, for example benzoyl-phenylalanine (BPA), which reacts with any nearby C-H bonds upon activation with UV light, and can be incorporated into the protein through usage of Amber stop codons and specialised tRNAs[182]. Until recently the only possible outcome of cross-linking experiments would be identification of interacting proteins by tandem mass spectrometry, without any spatial or sequence-specific information[183]. However, with the development of more sensitive mass spectrometers, crosslink enrichment and novel identification methods, it has become possible to identify the crosslinking sites on both proteins down to a tryptic peptide and sometimes even a single amino-acid (depending on the crosslinkers)[184]. Amine reactive NHS esters, which are some of the most commonly used crosslinkers for probing protein-protein interactions, are discussed in more detail in the following section.

4.1.2 Lysine-specific crosslinking

N-hydroxysuccinimide (NHS) esters are highly reactive towards primary amines[185], which includes the positively charged amino groups of lysine residues. Lysine is a common amino acid with an average relative occurrence in proteins (~5.5 ± 1.6 % in *E.coli*[186]), making it suitable for probing protein-protein interactions. The length of currently available crosslinkers with NHS-ester functional groups varies from 4.2 (TSAT) to 35.8 (BS(PEG)9) Å. Regular NHS-esters are non-soluble in water and require the use of an organic solvent such as DMSO, however sulfonated water soluble forms of these linkers also exist for instances where the use of DMSO is not possible and a water soluble crosslinker is required. BS3 and BS2G sulfo-NHS crosslinkers 11.4 Å and 7.7 Å long (see figure 4.1.2) were used in this study.

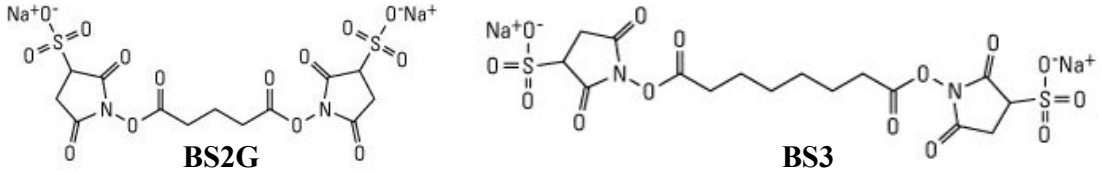


Figure 4.1.2 Sulfo-NHS crosslinkers: BS2 (7.7 Å) and BS3 (11.4 Å).

Apart from crosslinking to lysines, NHS-esters can also react with serine, tyrosine, and threonine residues as well as with protein N-termini[187]. At pH 6.0 they are reported to react preferentially with hydroxyl groups of tyrosine residues and protein N-termini, while at pH >8, they are known to preferentially interact with amine groups of lysine residues as well as the protein N-termini[187]. As well as participating in cross-linking, NHS esters also undergo two competing reactions: hydrolysis and amidation[187], which decrease the efficiency of cross-linking and need to be taken into account during data analysis. It is worth noting that in most common crosslinking conditions (in amine-free buffer pH 7-9) the preferred competing reaction is hydrolysis, and not amidation due to lack of substrate and unfavourable conditions for amidation. While NHS esters are certainly useful and

commonly used for crosslinking studies, there is one potentially serious issue that needs to be considered. Decreased trypsin digestion efficiency due to modification of lysine residues with the cross-linker may result in very long, highly charged and difficult to identify peptides, complicating the data analysis and cross-link identification.

4.1.3 Unambiguous identification of crosslinked peptides and residues using tandem mass spectrometry

Lysine specific crosslinking yields four types of peptides: common (no linker), mono-linked (linker bound, but not crosslinked to another peptide), loop-linked (linker crosslinks two residues on the same peptide) and inter-linked peptides (linker bound and crosslinked). Of these, common peptides are the most abundant in the sample, followed by mono-linked, inter-linked and loop-linked peptides, while the most useful are the rare inter-linked peptides. When using Q Exactive mass spectrometers, the top 10 protocol is frequently implemented: only top 10 ions from each primary scan (MS1) are routed for fragmentation and sequencing (MS2). Therefore, low abundance inter-linked peptides would be likely to get lost in the midst of the common peptides. To increase the number of unambiguously identified crosslinked peptides, an extra enrichment step, or selective sequencing should be implemented.

To enrich the crosslinked peptides one can purify them by either off-line or on-line size exclusion, or strong cation exchange chromatography, or by using a biotin-tagged crosslinker to purify and enrich the crosslinked peptides through streptavidin affinity chromatography[179]. Size exclusion chromatography would of course purify higher molecular weight products, which are more likely to contain crosslinked peptides. Trypsin cleaves the proteins at arginine and lysine residues, generating positively charged peptides.

This charge would double in two peptides, which are crosslinked to each other, strong cation exchange chromatography would help to separate crosslinked peptides from common ones before they are loaded onto the reverse phase column[188,189]. However, it is also possible to increase the chances of crosslink identifications without purifying the sample before the reverse phase chromatography. Again, due to the increased charge state of crosslinked peptides after electrospray ionization, they can be sequenced preferentially over the more common single peptide ions by setting a charge threshold. This way all parent ions would be scanned, while only ions with a charge state $\geq 3^+$ would be selected for MS2 sequencing event, so a higher proportion of the crosslinked peptides could be sequenced[179].

Each method has its advantages and disadvantages: while size exclusion and strong cation exchange chromatography may allow for more sensitivity towards low abundance cross-linked peptides, the additional purification step may lead to loss of material. The loss of material during the additional purification step would not be a concern for the selective fragmentation method, however its sensitivity is likely to be lower than that of the chromatography based approaches. Due to relatively low complexity of the samples (only four previously purified proteins), in this study we used the method of selective fragmentation of ions with charge state $\geq 3^+$ to increase the likelihood of crosslink identification.

Once the crosslinked peptides are isolated and fragmented, the fragment ion spectra need to be matched to theoretical fragment ion spectra generated from the database to determine their sequence. In the simplest scenario, the identification software generates a database, which contains tryptic peptides of all proteins in the search database, plus decoy peptides. Theoretical spectra for each of the peptides in the database are then compared to

experimental spectra. For linear peptides, one experimental spectrum could only match one theoretical spectrum, which allows search engines to reduce the number of comparison events by only comparing experimental spectra to theoretical spectra of peptides with the same precursor ion mass. However, each fragment ion spectrum of crosslinked peptides would contain information about both peptides, so each experimental spectrum would need to be compared to a combination of two theoretical spectra without being able to reduce the number of comparisons by precursor ion mass. This would dramatically increase the search space, as the number of possible combinations of spectra will be proportionate to n^2, where n is the number of tryptic peptides in the original database. This way if the *E. coli* proteome consists of 4000 proteins, which on average consist of 28 tryptic peptides, the number of possible theoretical spectra would increase from 112,000 to $6.27x10^9$, and this does not account for any missed cleavages or possible modifications and any reductions in search space made possible by search algorithms. At this point, not only would the search require an enormous amount of computing power, but also the chances of random positive identifications would increase increasing the false discovery rate (FDR). However, there are ways around this issue: one is using smaller databases composed of several proteins instead of several thousands (only possible if the constituents of the sample are known, and can result in increase in false positive identifications and incomplete biased data), and the preferred approach is using the software which relies on algorithms to reduce the search space.

4.1.4 Experimental design

The colicin E9 translocon was assembled and purified by Nick Housden (Kleanthous lab) as described in 2013 Housden et al.[33], with the following alterations: colicin E9 TR construct with two additional mutations in the T domain (A85C, and A104C to prevent

spontaneous unfolding) was used instead of the full-length colicin E9. The translocon was crosslinked and the products of the reaction separated on Blue Native PAGE and SDS-PAGE prior to trypsin digestion. The crosslinked tryptic peptides were then loaded onto a reverse phase chromatography column and analysed by tandem mass spectrometry. The detailed description of the methods can be found in sections 2.5-2.6. The components present in the samples were determined by MaxQuant[142]. While pLink[143,190] was used to identify the crosslinks using a small database composed of sequences of four most abundant proteins identified by MaxQuant (all four are components of the translocon). The search parameter for both searches can be found in section 2.7. The results were analysed using Xlink Analyzer (Chimera plug-in), PyMol, and HADDOCK 2.2 to validate the crosslinks and construct possible translocon models (see section 2.8).

4.2 Results

4.2.1 Optimisation of the lysine-specific crosslinking protocol for purified translocon

Because the colicin E9 translocon is thought to be a dynamic structure, in this experiment it was important to identify only the most easily formed crosslinks between the interacting proteins upon formation of the translocon. This means avoiding artifacts produced by over-crosslinking in solution and cross-links between the residues which are not close in the translocon but may move closer to each other during subsequent structural rearrangements. These issues were addressed by introducing a disulfide bridge between residues 85 and 104 in the T-domain of colicin E9, which prevents the initiation of global rearrangement during translocation[74], and using the lowest possible concentration of cross-linkers for the shortest possible period of time. Although there are established protocols for efficient crosslinking of soluble protein complexes with NHS ester crosslinkers (Thermo Scientific), they require optimization for each complex.

First, the optimum concentration of the BS3 reagent for crosslinking the translocon at 8 μM was determined by using concentrations within the recommended range: 0.5 – 2.5 mM, and at vast excess as a control (12.5 mM). The reactions were allowed to proceed for 30 minutes before they were quenched with 20 mM Tris pH 7.5. The samples were then boiled and loaded onto a 12% SDS-PAGE, the stained gel is shown in the figure 4.3.1 A. The sample cross-linked with the 12.5 mM BS3 was aggregated so the overall protein concentration in this sample appeared to be lower than in the rest. As for the other samples, there seems to be little difference in the amount of the cross-linked product achieved with different concentrations of the crosslinker, except for the sample with the 2.5 mM crosslinker, where a higher fraction of the sample failed to enter the resolving gel.

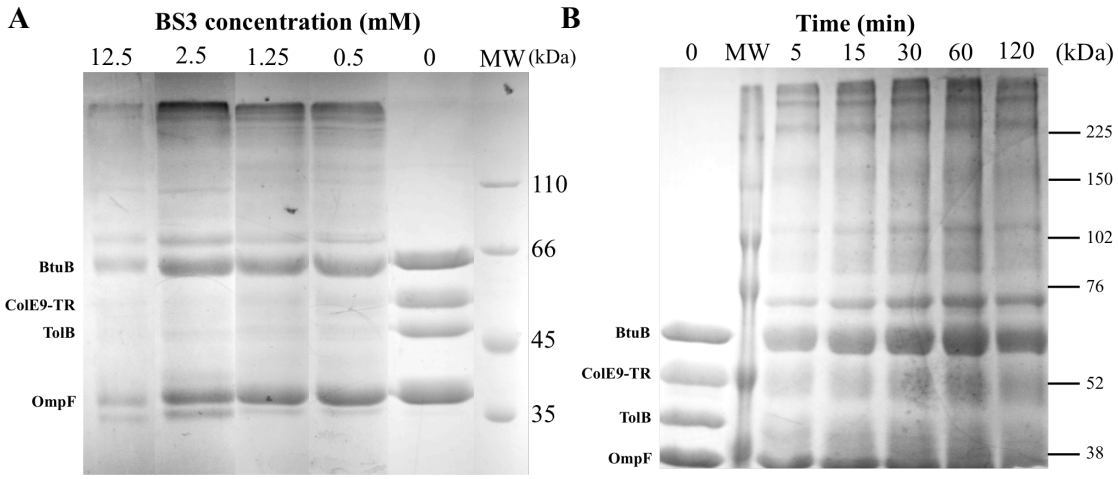


Figure 4.2.1 A. 12% SDS-PAGE of the products of the colicin E9 translocon crosslinking reaction with the range of concentrations of BS3 crosslinker. **B.** 10% SDS-PAGE of the products of the cross-linking reaction with 0.5 mM BS3 at different time points in course of 2 hours. Wide range molecular weight markers were used for this gel to gauge the size of the cross-linked products.

Because the lowest concentration of BS3 tested (0.5 mM) was enough to generate cross-linked products, this concentration was used for further experiments. Next, the influence of reaction time on formation of the cross-linked products was tested. The reaction was set up using the same concentration of the translocon as before and 0.5 mM concentration of the BS3 cross-linker. Samples were taken at 5, 15, 30, 60, and 120 minutes, quenched with 20 mM Tris pH 7.5 and subsequently analysed on 10% SDS-PAGE gel shown in 4.2.1 B. It appears that even after 5 minutes, the crosslinking reaction was complete. This result was expected, as the NHS esters are unstable in aqueous environment and undergo rapid hydrolysis. This reaction competes with the crosslinking and therefore makes excessive crosslinking due to random contacts unlikely. As a result of these tests all cross-linking reactions were conducted with 0.5 mM concentration of crosslinker for 5 minutes before quenching with Tris to a final concentration of 20 mM. 10% acrylamide gels and wide range molecular markers were used for all subsequent SDS-PAGE experiments.

4.2.2 Crosslinking the translocon with BS2 and BS3 crosslinkers

The crosslinking procedure was carried out as described in the previous section, however there were two issues with the bands pattern on the SDS-PAGE shown in figure 4.2.1: some proteins were too large to enter the resolving gel, and most of the proteins within the complex did not cross-link to their interaction partners. This posed the risk of identifying some inter-translocon cross-links and some intra-protein links from the proteins, which were not part of the assembled translocon during crosslinking. A simple way around this issue was to separate the crosslinked and non-crosslinked translocon from the rest of the products on the Blue Native (BN) PAGE prior to the SDS-PAGE separation, after which the relevant bands were digested with trypsin and the peptides analysed using in-house LC-MS/MS system as described in section 2.6. Both, the 1st dimension BN PAGE and 2nd dimension SDS-PAGE gels are shown in the figure 4.2.2

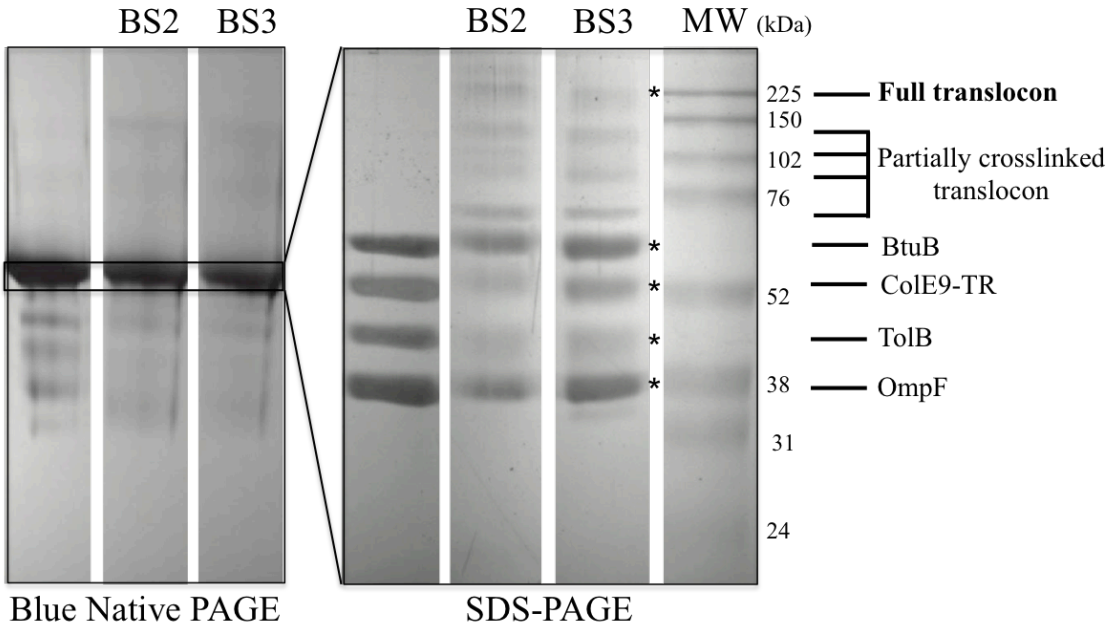


Figure 4.2.2 Stained 1st dimension BN PAGE and 2nd dimension SDS-PAGE of the crosslinked and non-crosslinked samples. The proteins in each band were identified by tandem mass spectrometry and are labelled on the right side of the gel. The full translocon band likely includes all four proteins cross-linked in 1:1:1:1 ratio. Hence only one of three copies of OmpF in each translocon was cross-linked to colicin E9. * - Bands, which were used to identify crosslinks discussed in this chapter.

To identify the proteins in each major band, the data from the LC-MS/MS experiments were searched using MaxQuant against the *E. coli* proteome database as well as the potential contaminants to identify proteins present in each band[142]. The approximate stoichiometry was determined by comparing the intensity values (the sum of intensity values of three highest abundance peptides from each protein[142]). Only the highest molecular weight (MW) band in each of the lanes was found to contain all four proteins in high abundance, while other lower MW bands contained varying amounts of all four proteins making it impossible to unambiguously identify which proteins were responsible for the staining. Interestingly, the intact translocon was expected to have one molecule of colicin E9, one molecule of TolB, one molecule of BtuB, and one molecules of OmpF[33], so the expected molecular weight of the full translocon was 271 kDa, however as seen on the gel in figure 4.2.2 the highest molecular weight band is just under or around the 225 kDa mark. This means that only 1-2 of the 3 OmpF molecules in the translocon were crosslinked to the rest of the proteins. This was expected as low crosslinking efficiency makes it highly unlikely that five particular inter-protein cross-links would occur between six proteins connecting them to each other within the same translocon entity. The majority of the crosslinks were expected to occur within each protein, and not to the other interacting proteins. This is reflected in the figure 4.2.2 by the presence of a large proportion of single translocon components in the crosslinked sample on the SDS-PAGE gel.

4.2.3 Identification and validation of crosslinks

Each trypsin digested band from the gel shown in figure 4.2.2 generated approximately 23,000 MS2 spectra when sequenced by tandem mass spectrometry. These spectra were matched against theoretical ion spectra of peptides in a small database by pLink. The

average parent ion mass tolerance tolerance was set at 5Da to include all monoisotopic and isotopic peaks for improved assignment of the monoisotopic peaks to the spectra[190]. However, ±5 Da is a large window that may lead to ambiguities during spectrum matching, to avoid this isotopic precursor ion tolerance was set at ± 10 ppm[190]. The fragment ion tolerance was set at 20 ppm. To increase the sensitivity of the search, the search database was constructed from results obtained from the identification search with MaxQuant and included all four high-abundance proteins, which are known to be part of the translocon. Although other bacterial proteins were also identified by MaxQuant their intensity values were several orders of magnitude lower than those of the translocon components. Therefore they were not included in the small search database. To ensure that these proteins were contaminants and did not participate in translocon formation, the spectra obtained from the full translocon sample were also searched against the theoretical spectra generated from the entire *E. coli* protein database. As expected, no cross-links to proteins other than those involved in formation of the translocon were identified within the imposed E-value cut-off.

While the bands produced by crosslinking with both BS2 and BS3 were analysed, the results largely overlapped, while the quality of the spectra and identifications from the BS2 crosslinking experiment was inferior to those obtained from BS3 crosslinking. Therefore, only data acquired from the BS3 experiments will be discussed from here on. Search results from samples crosslinked with BS3 linker at 5% false discovery rate (FDR) are summarized in table 4.2.3.

Table 4.2.3 Number of peptide spectrum matches (PSMs) from samples crosslinked with BS3

PSMs	Common	Loop-linked	Mono-linked	Inter-linked
OmpF	181	0	23	15
TolB	1515	26	195	79
ColE9	1879	34	218	316
BtuB	2016	23	366	900
Full-translocon	1744	25	239	345

The low number of common peptide sequence identifications in all samples compared to the total number of spectra (<10% in all cases) is likely due to the use of selective sequencing of highly charged peptides, which lead to loss of data on the most common doubly charged peptides, but increased the chances of identifying cross-linked peptides.

At 5% FDR several identified spectra lacked sequence information, so a cut-off point was set for the identification E-values: any identified crosslinks from monomeric protein bands with E-values $\geq 10^{-4}$ were discarded. For the cross-linked translocon, the cut-off value was set at $5x10^{-4}$, because in this sample the identification E-values were higher overall due to the increased complexity, and decreased amount of material. The crosslinks that were only identified by a single spectrum were discarded also. This yielded identification of eight inter-protein cross-links and 15 intra-protein cross-links in the full-translocon band, and 14 intra-BtuB links, 24 intra-ColE9-TR links, one intra-TolB link, and one intra-OmpF link in respective monomeric protein bands. Unlike the structure of the whole translocon, crystal structures of its individual components have been solved, making the intra-protein cross-links ideal for validation of the crosslinking approach. Using XlinkAnalyzer, all crosslinks found in each monomeric protein band were mapped onto the appropriate crystal structure to evaluate the distances between crosslinked lysines in crystal structures against the 11.4 Å restraint imposed by the BS3 linker, the results are summarised in figure 4.2.3

Each of the crosslinks identified within OmpF and TolB agrees with existing crystal structures: the linked lysines are perfectly positioned in close proximity with their amine groups facing each other. The intra-protein crosslinks within BtuB do not agree quite as well. While the majority of the crosslinks were identified between the lysines that are close to each other, four pairs of lysines are positioned further away (the amine groups are not close enough to be connected by an 11.4 Å linker). Of these four pairs, however, three have

at least one lysine in a long and flexible extra-cellular loop, which are likely to assume a variety of conformations not represented in the crystal structure. Therefore only one crosslink is left, and it is unlikely to occur in monomeric BtuB as it links two lysines on opposite sides of the β-barrel. Although there are only two spectra that can be matched to this crosslink, they contain enough information to make this identification reliable. While this crosslink could not have occurred in a monomeric protein, it could occur in a BtuB dimer (a potential contaminant in the sample). Overall, the crosslinks identified in OmpF, TolB and BtuB are reasonable and attest to the applicability of the used method for identifying crosslinks and reliability of the findings. However, as shown in figure 4.2.3, the majority of crosslinks identified in colicin E9 monomer are not compatible with the known crystal structure.

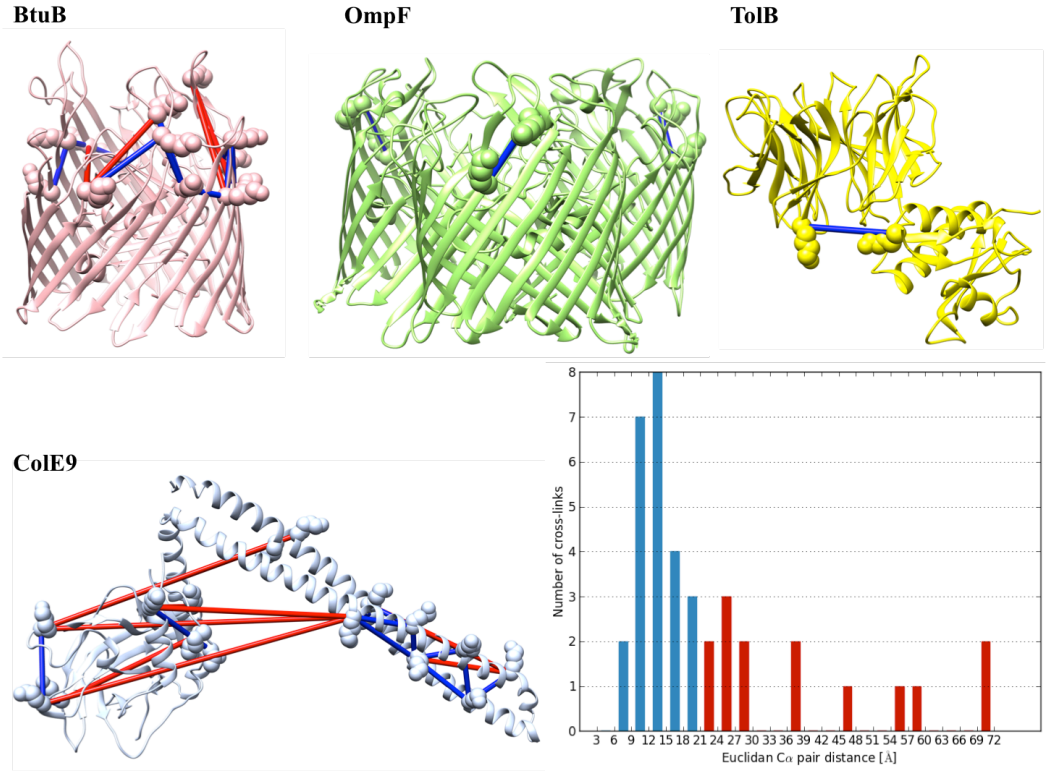


Figure 4.2.3 Intra-protein crosslinks shown on crystal structures of individual proteins. Distances between Cα ≤20 Å are shown in blue and are thought to be satisfied, while distances >20 Å are shown in red and are thought to be incompatible with the crystal structures. In the bottom right is the distribution of distances between Cα of crosslinked lysines shown in this figure.

This discrepancy between the crystal structure of the colicin and the crosslinking data was unexpected, suggesting that the T and R domains may be required to move relative to each other during the translocon formation. Because these crosslinks were observed when the colicin was already part of the purified translocon, I have investigated whether or not they occur in monomeric colicin E9, and in colicin E9 bound to BtuB. This would help to determine whether the crosslinks are a result of conformational dynamics of colicin E9 in solution, or a result of a conformational change due to the binding of colicin E9 R domain to BtuB, or the binding of T domain to OmpF and TolB during translocon assembly.

4.2.4 Probing the flexibility of colicin E9 outside the translocon

Given the results described in the previous section, and the known requirement for freedom of movement of the helices of the coiled coil relative to each other, it is likely that colicin E9 may be inherently more flexible than previously thought. More specifically, the 60° angle between the T and R domains of the colicin observed in both crystal structures of colicin E9 and E3 may be variable in solution as in the crystal this angle is stabilised by the tip of the second colicin molecule as shown in figure 4.2.4 A. Therefore it is possible that when the protein is a monomer, this angle is dynamic and acts as a hinge, allowing greater freedom for the T-domain to probe the surrounding area for its translocator protein OmpF. The conformations of colicin E9 in solution on its own, and in complex with BtuB were investigated through identifying pairs of neighbouring lysines in each condition, and comparing the results to crosslinks found in colicin E9 within the translocon. The crosslinking reactions were performed in the same conditions as the translocon crosslinking using purified colicin E9-TR, and a 1:1 stoichiometric mixture between colicin E9-TR and BtuB. Because we were interested in sampling all possible intra-protein cross-links that occur in solution, there was no need for separating the products of the reaction on BN

PAGE, therefore all reaction products were analysed using 10% acrylamide SDS-PAGE, which is shown in figure 4.2.4 B.

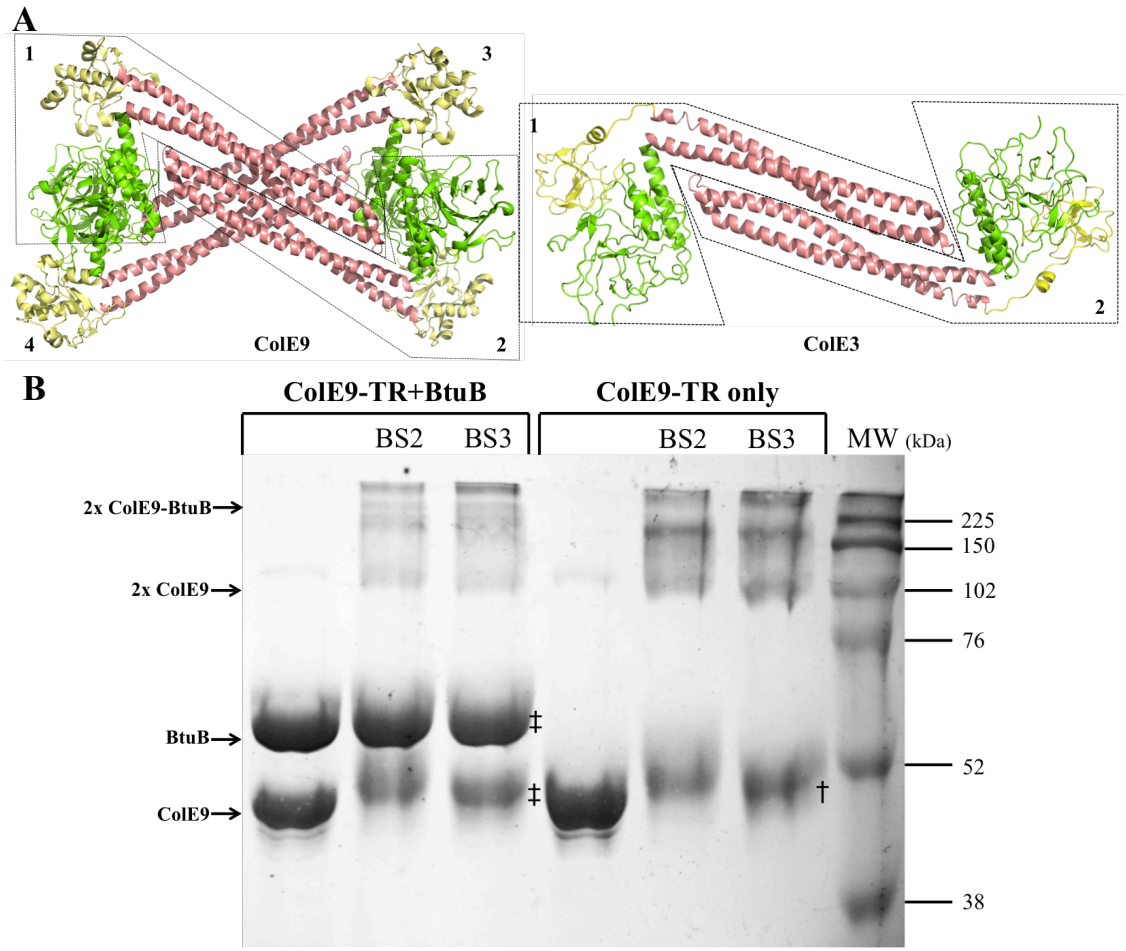


Figure 4.2.4 A. Crystal structures of colicins E9 and E3 showing the entire asymmetric units. PDB ID: 5EW5, 1JCH. N-terminal T domains are shown in green, R domains in coral, and the cytotoxic domains in yellow. Individual colicin molecules are numbered and framed for clarity. **B.** Stained SDS-PAGE of crosslinked and non-crosslinked samples of colicin E9-TR in aqueous solution, as well as colicin E9-TR bound to BtuB in β-Octyl glucosode. As with the previous gel, the protein components of the bands and stoichiometry were identified by tandem mass spectrometry and considering the molecular weight of the cross-linked complexes the number of proteins involved were estimated.
†, ‡ - Bands used for crosslink identification.

The band patterns generated by the crosslinked samples of colicin E9 and a colicin E9-BtuB complex were unexpected, as the only difference between them was the presence of a band with molecular weight over 225 kDa in the colicin E9-BtuB sample. This would correspond to twice the expected molecular weight of ColE9-TR-BtuB complex, suggesting that colicin E9-BtuB complex has dimerised.

4.2.5 Changes in conformation of colicin E9 when it is part of the translocon

The mass spectrometry data collected for monomeric colicin E9 bands from both samples were analysed in pLink using the same search database and mass tolerance parameters as described above. The results at 5% FDR were filtered according to the cut-off E-value set at $5x10^{-4}$ and summarised in table 4.2.5

Table 4.2.5 Intra-protein cross-links identified in various monomeric colicin E9 bands with the corresponding E-values: colicin E9-TR in solution, colicin E9-TR from the BtuB complex, colicin E9-TR from the translocon. The indication 'not found' may include identified crosslinks with E-values above the imposed threshold. Crosslinks, which cannot be satisfied within a monomeric protein, but are possible within a tetrameric complex, are highlighted red. These peptides were likely carried over from the previous mass spectrometry experiment (crosslink identification of the highest molecular weight band from the same well on the gel).

	E-values of best identified spectra		
Cross-linked residues	ColE9 alone	ColE9 from ColE9-BtuB complex	ColE9 from translocon
K125-S2	2.70E-07	3.39E-07	not found
K177-K277	2.13E-10	3.05E-07	5.38E-17
K177-K347	not found	1.32E-12	2.99E-10
K177-K434	not found	not found	1.27E-07
K191-K125	not found	not found	4.41E-08
K191-K436	7.65E-08	1.28E-07	1.19E-09
K213-K191	not found	6.74E-06	not found
K255-K213	2.16E-05	1.85E-04	1.07E-05
K277-K213	not found	not found	4.02E-10
K277-K347	not found	not found	2.13E-15
K298-K436	not found	not found	8.67E-05
K347-S2	2.59E-09	not found	not found
K347-K125	not found	1.20E-05	1.49E-08
K347-K347	not found	7.42E-10	not found
K347-K356	3.11E-16	2.57E-14	3.19E-14
K347-K417	1.90E-08	not found	1.55E-07
K356-K419	1.86E-16	1.16E-06	1.02E-15
K364-K397	not found	not found	2.74E-08
K364-K408	2.49E-10	not found	2.60E-14
K364-K419	4.43E-11	not found	1.69E-13
K374-S2	1.06E-06	not found	not found
K374-K347	not found	1.31E-08	not found
K374-K356	6.82E-14	1.15E-09	7.46E-11
K374-K397	7.78E-11	1.27E-07	1.54E-06
K397-S2	8.61E-06	not found	not found
K397-K347	5.43E-15	2.13E-12	not found
K397-K356	2.17E-05	not found	not found
K408-S2	1.79E-05	5.52E-08	not found
K408-K347	5.03E-14	1.98E-16	3.91E-17
K408-K356	not found	not found	3.89E-08
K408-K374	2.00E-07	not found	1.05E-10
K408-K397	9.96E-15	1.55E-11	1.96E-10
K408-K419	1.63E-15	2.96E-07	not found
K417-S2	3.98E-06	not found	not found
K417-K347	1.79E-14	3.07E-15	9.38E-18
K417-K356	3.46E-12	3.03E-10	3.57E-14
K417-K364	not found	3.17E-04	not found
K417-K397	6.13E-05	not found	not found
K419-S2	3.64E-05	not found	not found
K419-K347	8.67E-14	9.94E-11	7.39E-14
K419-K397	3.46E-07	not found	not found
K419-K419	not found	1.26E-05	not found
K436-S2	2.88E-06	not found	not found

The crosslinks, which are listed in table 4.2.5, were analysed using XlinkAnalyzer plug-in within Chimera, and the results are shown in figure 4.2.5. 11 crosslinks were identified in all three of the samples and the relative positions of the lysines involved were treated as constant during the process of translocation. A large fraction of identified crosslinks was found to be unique to each protein sample: nine crosslinks in colicin E9 on its own, five in the colicin E9-BtuB sample, and seven were unique to colicin E9 molecules which were part of the translocon. These crosslinks are thought to occur due to the differences in conformation and flexibility of the colicin E9 unique to particular binding partnerships. The crosslinks identified in two of the three samples also arise from the differences in conformation and flexibility of the colicin, but because they are not unique to one condition, they help to establish the order of the conformational changes relative to ColE9-BtuB and ColE9-OmpF binding events.

Eight of unique crosslinks identified in single colicin E9 in solution (band marked † in fig. 4.2.4 B) are between lysines in the structured T-R domains of the colicin and its disordered N-terminus. Two of these cross-links are also found in colicin E9 from the BtuB-ColE9-TR complex (band ‡ in fig. 4.2.4 B), suggesting that the flexibility of the N-terminus is greatest when the colicin is in solution and decreases when the protein binds to BtuB. These crosslinks were not found in the colicin isolated from the translocon sample, because the N-terminus of the colicin is separated from the rest of the protein by TolB and OmpF. In the *in vivo* assembled translocon the disordered N-terminus interacts with the lumen of OmpF from the periplasmic side and is thus separated from the rest of the protein by the outer membrane. This separation should be preserved throughout purification. The rest of the unique crosslinks are between the helices of the coiled coil region as shown in figure 4.2.5. Their absence in the other colicin E9 samples is likely due to decreased flexibility of the molecule caused by binding of the colicin to other proteins.

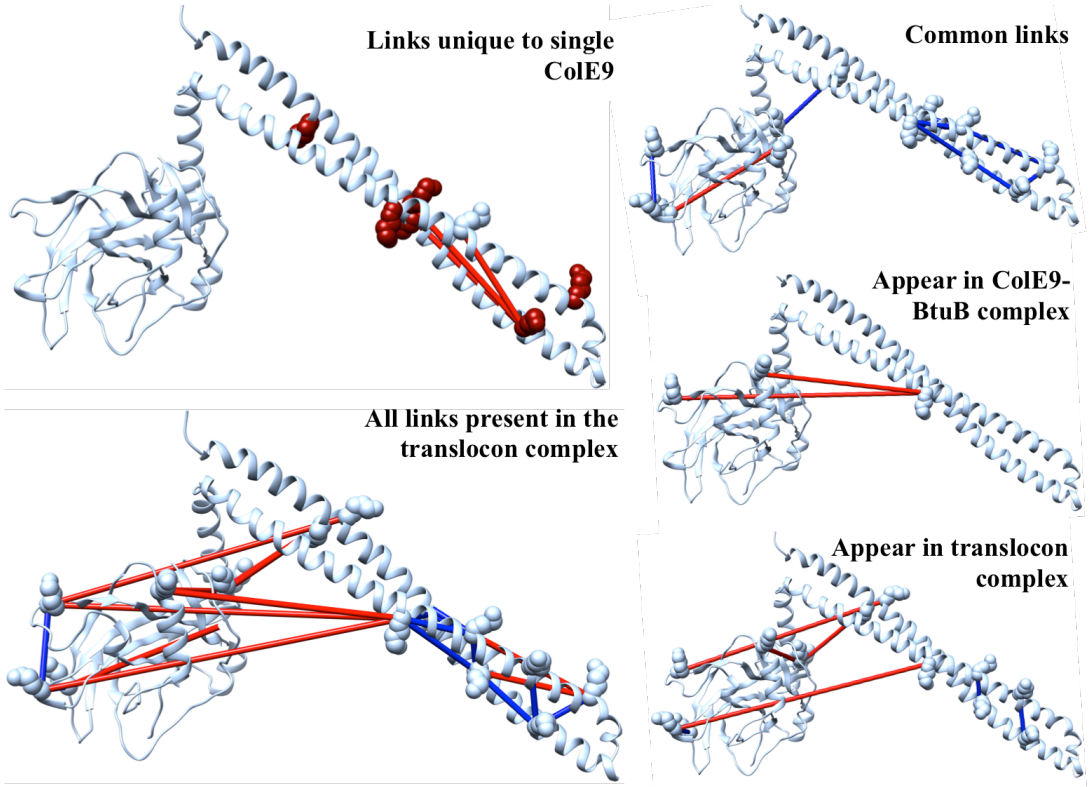


Figure 4.2.5 Summary of crosslinks identified in colicin E9 mapped onto the crystal structure of the colicin E9-TR construct (PDB ID: 5EW5). Clock-wise: from unique crosslinks found in monomeric colicin E9 in solution to the crosslinks identified in the colicin E9 that was part to the translocon, highlighting the experiments at which the crosslinks appear for the first time. Unique cross-links found in ColE9-BtuB complex are not shown. Distances between Cα ≤20 Å are shown in blue and are thought to be satisfied, while distances >20 Å are shown in red and are thought to be incompatible with the crystal structures. Lysines, which are cross-linked to the N-terminus of colicin E9 are coloured burgundy.

The presence of crosslinks unique to the ColE9-BtuB complex was initially surprising, however upon further examination it became apparent that none of the crosslinks could possibly occur in the monomeric form of colicin E9. In fact, all but one of these crosslinks (highlighted in red in table 4.2.5) could be satisfied if the protein was in its tetrameric form (as shown in the crystal structure in figure 4.2.4 A), which was also present in the highest molecular weight band from the same experiment. Because the peptides from that band were sequenced immediately prior to those from the band that contained monomeric colicin E9, the cross-linked peptides are likely to have come from the residual material from the previous run. The origin of the K213-K191 crosslink remains unclear, as it does not fit the

crystal structure, nor it appears in any of the other bands and experiments, which means that it is likely to simply be a false identification.

As can be seen from the figure and table 4.2.5, aside from the crosslinks common to all samples, all but two crosslinks found in colicin E9 from the translocon are absent from the ColE9-BtuB and colicin E9 samples, suggesting that the biggest conformational changes in colicin E9 are due to its interactions with OmpF and TolB, and not its association with BtuB. However, at least some structural rearrangements in colicin E9 are likely to be a consequence of binding to BtuB as evidenced by the two crosslinks between the T and R domains.

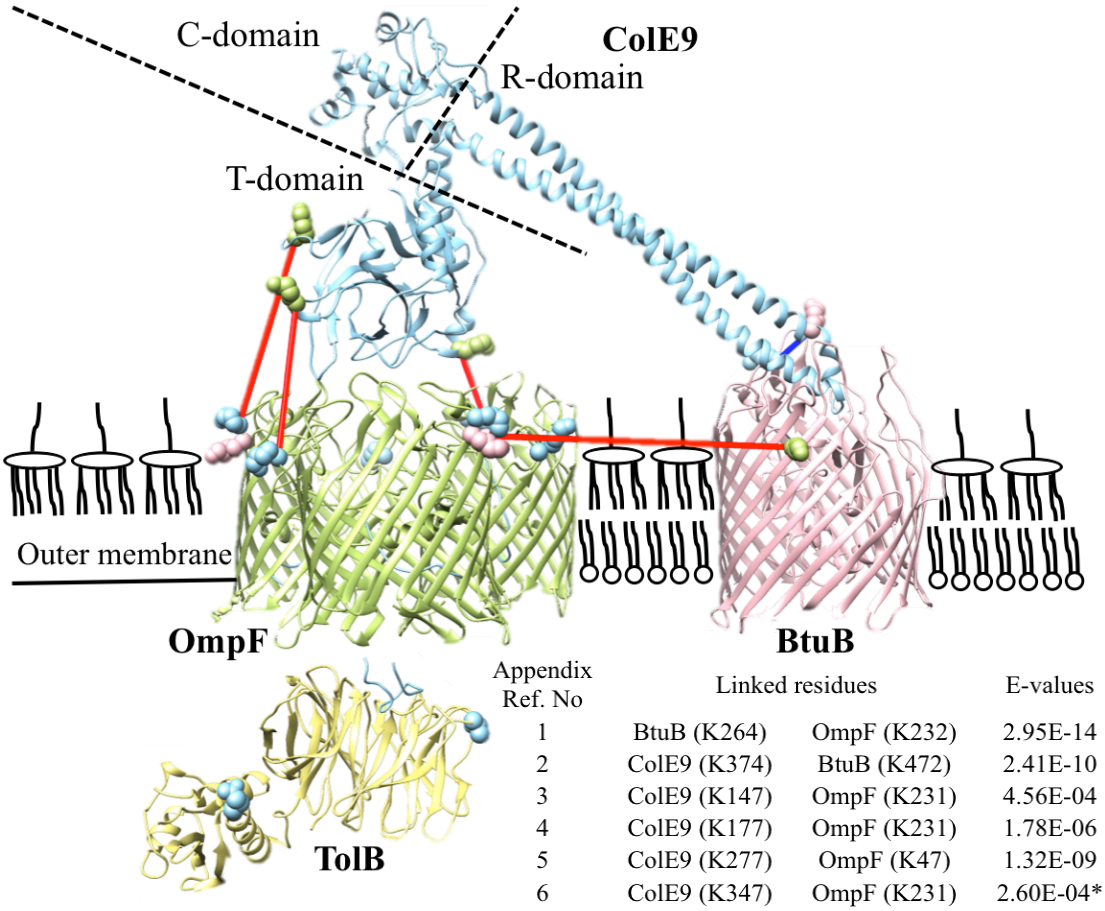
4.2.6 Mapping intermolecular crosslinks to the current translocon

While intra-molecular crosslinks were identified from the monomeric protein bands, the high molecular weight band of the full translocon (figure 4.2.2) was used to identify any inter-molecular crosslinks. The small database search of spectra obtained from the digested full translocon band resulted in 2353 peptide spectrum matches (PSMs): 1744 to common peptides (non-modified), 25 to loop-linked peptides, 239 to mono-linked peptides and 345 to cross-linked peptides at 5% FDR. This corresponds to 124 unique common peptides, 4 unique loop linked peptides, 32 unique mono-linked peptides and 57 unique cross-linked peptides.

The translocon contains a total of 117 lysines, of which 58 are accessible to the crosslinkers (found to be either mono-, loop-, or inter-linked), of which in turn 33 were inter-linked with one or more other lysines. Although 5% FDR was applied to filter the results, due to the lack of fragment ion information in some spectra with higher E-values, an extra threshold

was imposed on the E-values: $5x10^{-4}$. After filtering the results based on this principle, 22 unique cross-links were identified. Of these, 14 were intra-protein crosslinks: four within colicin E9, one within OmpF, and one within BtuB. Eight inter-protein cross-links were also identified: four between colicin E9 and OmpF, one between colicin E9 and BtuB, two between colicin E9 and TolB, and one between OmpF and BtuB. While all intra-protein crosslink spectra within the applied threshold were of good quality and matched those already identified, upon closer examination one of the inter-protein spectra appeared to be of unsuitable quality for reliable identification. Its identification E-value was within the applied threshold, however it appears that the E-value was skewed by the fact that the spectrum had a lot of information about the α-peptide, while almost none for the β-peptide. This spectrum along with the other spectra identifying inter-protein cross-links is shown in the appendix (figure I), while the crosslink it identified was not included in further analysis.

The inter-protein cross-links were visualised with XlinkAnalyzer on the current model of the colicin E9 translocon as proposed by Housden et al.[33] The crystal structures of OmpF, BtuB and TolB in complex with different regions of colicin E9 were used to align the proteins as shown in figure 4.2.6. Of seven inter-protein crosslinks, only one, between colicin E9 and BtuB, was found to be in accordance to this model. The cross-links between the N-terminus of the colicin E9 and TolB contradict the model where the N-terminus is inserted into the lumen of OmpF, however this may be simply due to the effect of high pH at which the crosslinking reaction was performed (the insertion in known to occur at pH 6.5-6.7, compared to 7.4 of the crosslinking reaction[33]). The interactions of OmpF with BtuB, and the folded part of the T domain of ColE9 with OmpF were not previously observed so the identified cross-links could not be evaluated against the known information. The overall result of mapping the crosslinks to the crystal structures suggests that the translocon structure is more compact than the current model suggests.



Appendix Ref. No	Linked residues		E-values
1	BtuB (K264)	OmpF (K232)	2.95E-14
2	ColE9 (K374)	BtuB (K472)	2.41E-10
3	ColE9 (K147)	OmpF (K231)	4.56E-04
4	ColE9 (K177)	OmpF (K231)	1.78E-06
5	ColE9 (K277)	OmpF (K47)	1.32E-09
6	ColE9 (K347)	OmpF (K231)	2.60E-04*
7	ColE9 (S2)	TolB (K117)	1.03E-09
8	ColE9 (S2)	TolB (243)	7.07E-06

Figure 4.2.6 Current translocon model with intermolecular links found in the translocon. Crosslinked lysines are coloured according to the subunit to which they are cross-linked. For simplicity, only the crosslinks to the closest OmpF monomer are shown in this figure. Red lines indicate cross-links, which do not agree with the current model because the Cα of the lysines are located ≥ 20 Å apart. One cross-link is marked blue as the distance between Cα of the two lysines is < 20 Å, hence it is consistent with the crystal structure and therefore with the current model. Lysines marked on TolB are linked to the N-terminus of the ColE9 (S2). Table of identified inter-protein crosslinks, with their corresponding identification E-values, is shown in the bottom-right corner. * E-value of this crosslink is within the applied threshold, but the spectrum lacks information on one of the peptides. Matched fragment ion spectra for these identifications can be found in the appendix.

4.2.7 Remodeling the translocon complex to fit the crosslinking data

The OmpF-ColE9-TR crosslinks were used as distance restraints to dock the colicin E9 molecule to the OmpF in HADDOCK 2.2[146]. Two lysines on the T domain of colicin (K177 and K147) located ~ 50Å apart were found to crosslink to the same residue of the OmpF (K231), which is only possible if these crosslinks are made to different OmpF monomers.

The third residue of the T domain (K277) ~20 and 52Å away from K177 and K147 respectively was found to cross-link to K47 of OmpF, which again is only possible if each of the crosslinked residues on the T-domain contacts a different OmpF monomer. Therefore the OmpF monomers and the lysine residues on the colicin E9-T domain were numbered clockwise and distance restraints of 12±5 Å were applied to each pair. Dockings using pairs of cross-links were also tested. The best model satisfying all distance restraints was achieved with the pair of crosslinks: ColE9(147)-OmpF(231) and ColE9(277)-OmpF(47). The best structure from this docking with a HADDOCK score of -105.0306 is shown in figure 4.2.7 A. The colicin fits well into the middle of the OmpF trimer with no clashes and a total calculated buried surface area of 2310 $Å^2$, leaving room for the 85 residues of the colicin E9 N-terminal domain, which are not present in the crystal structure. However, if the structure of the colicin remains consistent with the crystal structure, this arrangement puts the R-domain of the colicin E9 into a position that is incompatible with BtuB binding at the membrane. The colicin would have to bend at the hinge region between the T and R domains in order to be in contact with both OmpF and BtuB. The intra-protein crosslinks found in colicin E9 when it was part of the translocon support the need for this kind of movement.

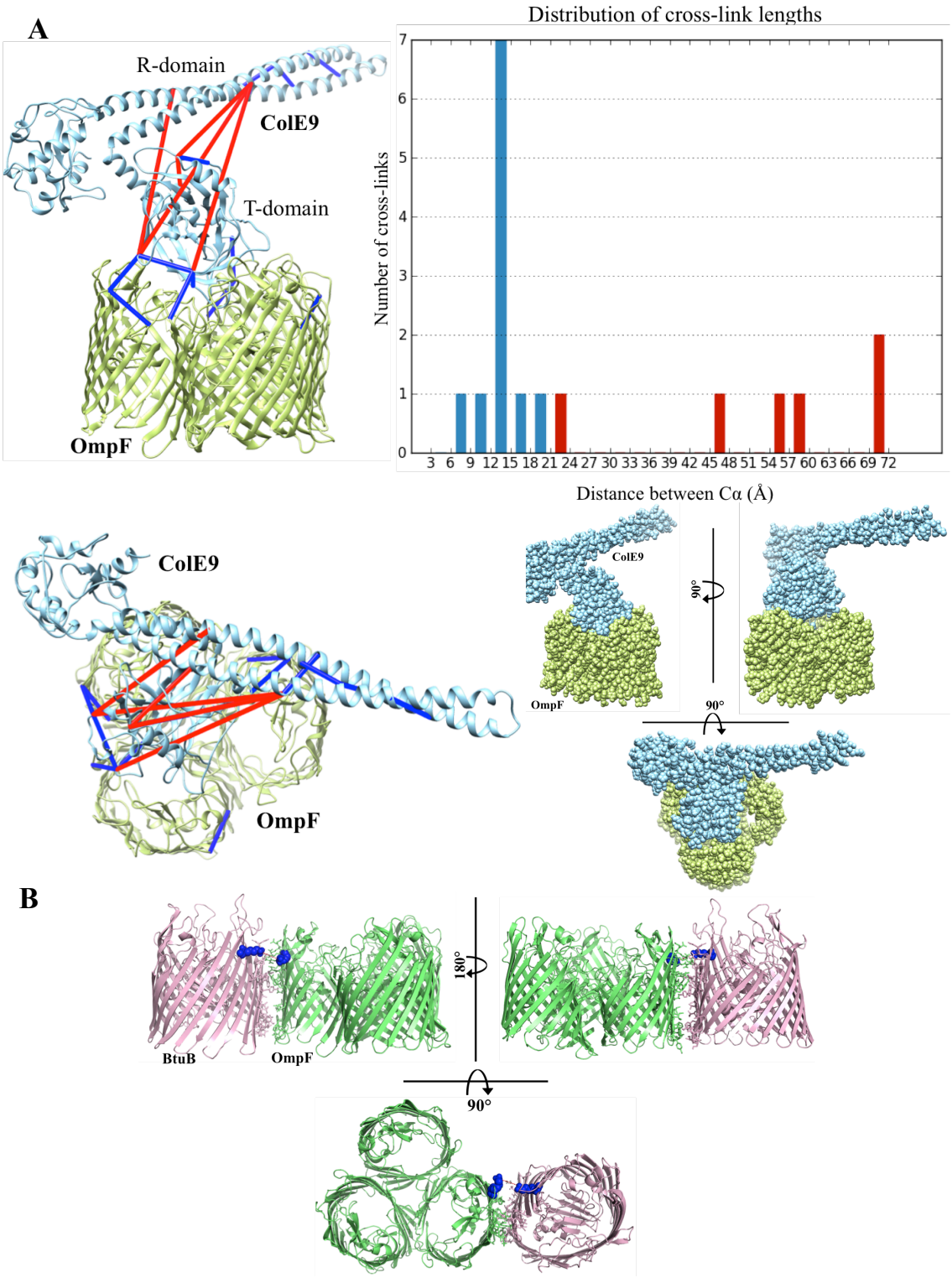


Figure 4.2.7 A. Best ColE9-OmpF docking result from HADDOCK with the histogram distances between Cα of cross-linked lysines. The crosslinks between lysines ≤ 20 Å away are shown with blue lines and the crosslinks between lysines >20 Å away from each other are shown with red lines. Bottom right is a space-fill view of the protein interface. **B.** Manually aligned membrane proteins BtuB and OmpF using a previously predicted interface as a guideline. The crosslinked lysine residues are shown in blue, with distance of 9.6 Å between the amine groups as measured in PyMol.

Recent studies suggest that the outer membrane proteins of *E. coli* are arranged into stable protein-rich islands and that BtuB and OmpF are capable of interacting with each other forming stable aggregates in lipid environment[141]. The crosslink observed between OmpF and BtuB as part of the assembled colicin E9 translocon is the first direct evidence that these proteins do indeed interact *in vivo* as the translocon was assembled and isolated from bacteria. Rassam et al., have predicted two OmpF-BtuB interaction interfaces (one side of OmpF can contact to different areas of BtuB) by molecular dynamics simulations[141]. One of these interfaces puts the lysines involved in the crosslink between BtuB and OmpF identified in this study in a perfect position for cross-linking. Because the interacting surfaces identified by MD simulations were ambiguous, there was only one unambiguous restraint derived from the crosslink, and the overall tendency of HADDOCK to cover the largest possible surface area, the results of docking with HADDOCK software were poor. Therefore the interaction interface was approximated manually in PyMol using the previously defined interface instead, and the crosslink was mapped onto the resulting structure shown in figure 4.2.7 B. The distance between the amino groups of lysines involved in crosslinking in this structure is 9.6 Å, which is just under the length of the BS3 crosslinker in extended conformation.

Crosslinks between colicin E9 and TolB were not used in this docking as they link TolB to the disordered N-terminus of the colicin, which is known to be inserted into the OmpF *in vivo*[33] and therefore would not be close to lysine residues on TolB. These crosslinks are likely to be artifactual due to the pH of the buffers used during crosslinking (section 4.2.6). Even if the crosslinks agreed with the other data, due to intrinsic disorder within the first 85 residues, they do not provide enough information about the interaction of the two proteins to model the interface between ColE9 and TolB.

Because the interfaces between each of the three proteins: colicin E9, OmpF and BtuB are now known, we can propose a model of colicin E9 translocon assembly at the outer membrane.

4.2.8 New model for the colicin E9 translocon

On the first attempt to propose a model for the translocon structure at the outer membrane, all three proteins were aligned according to the proposed interfaces: ColE9-OmpF according to the HADDOCK docking result, BtuB-ColE9 according to crystal structure, BtuB-OmpF according to the previously identified by MD interface. Two variants of this alignment are shown in figure 4.2.8 A.

Colicin interacts with three OmpF monomers, making the structure asymmetric and resulting in three possible structures where BtuB is bound to each of the monomers. One such structure (in contact with OmpF monomer 1) is unlikely to occur due to a significant mismatch between the length of the R domain of the colicin and the distance between the T-R hinge and the R-domain binding site on BtuB. This arrangement was excluded from further analysis leaving us with only two possible structures. However, both of them disagree with the crystal structure of colicin E3 R domain bound to BtuB, as even with movement within the hinge region, the lateral angle of ColE9-R binding to BtuB would change relative to that observed in the crystal structure by approximately 60-90°. Since most of the interactions between the R domain of colicin E9 with BtuB are made through the second helix of the coiled coil, the binding angle could be preserved by complete opening of coiled coil. However, this conformation would contradict previous structural studies[191] and the intra-coiled coil crosslinks identified in this study.

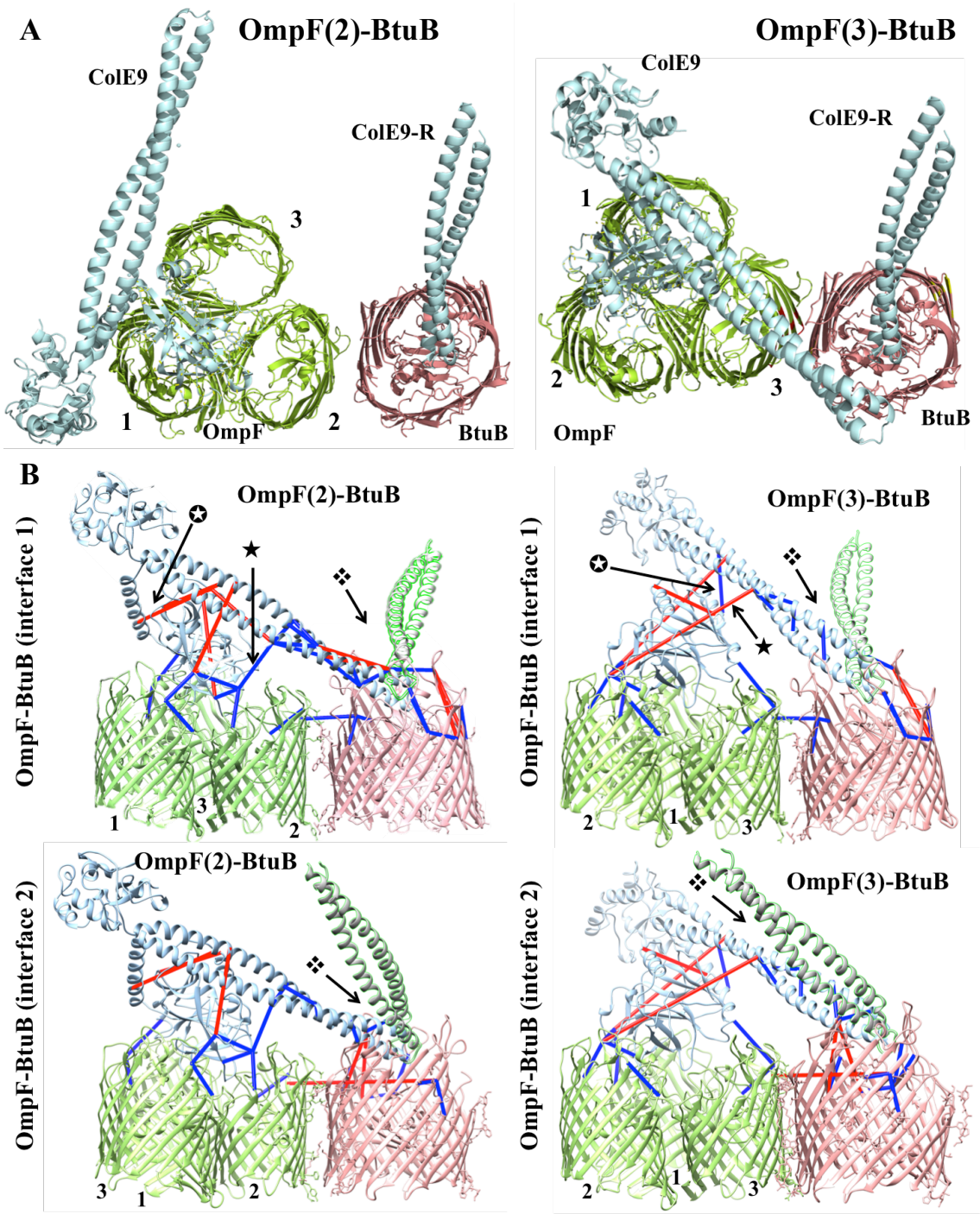


Figure 4.2.8 A. Two variants of BtuB positioning relative to the colicin E9-OmpF complex: BtuB contacts OmpF monomers 2 and 3. The position of the colicin E3 (E9) R domain relative to BtuB barrel, as seen in the crystal structure, is shown in light blue on top of the BtuB molecule (PDB ID: 1UJW). **B.** Possible models of the colicin E9 translocon with identified crosslinks shown according to the colour scheme used in previous figures. Top: variant I, and variant II of the translocon built according to the interaction surface and the identified crosslink. Bottom: variant I, and variant II of the translocon built according to the second OmpF-BtuB interface identified by MD. The crystal structure of the R-domain bound to BtuB is shown in white with green border for reference. Changing crosslink distances are shown with arrows. Different R domain angle compared to crystal structure is shown with ❖.

Another hypothesis, which would allow all of the predicted interfaces to fit together, would be the counter-clockwise rotation of the colicin R domain relative to BtuB in a 'ball-and-socket' fashion upon formation of the full translocon. If this rotation were possible, it would allow the predicted OmpF-ColE9 and OmpF-BtuB interfaces to occur simultaneously, while also satisfying some of the identified crosslinks between colicin E9 T and R domains. It appears impossible to satisfy all of the crosslinks between the T and R domains with any one conformation of colicin E9 suggesting that it is likely that there is more than one possible arrangement of proteins within the translocon. At this point, the second interface identified by MD simulations, although it is not supported by the crosslinking data, cannot be overlooked either, as when it is satisfied, no pair of lysines is suitably positioned for crosslinking with an 11.4 Å long BS3 linker. Therefore, at least four different options are plausible: variants I and II of translocons assembled via each of the OmpF-BtuB interfaces identified by Rassam et al. Examples of such translocons are shown in figure 4.2.8 B.

4.3 Discussion

4.3.1 The impact of these results on current understanding of ColE9 recognition and translocation

We know that colicin E9 initiates translocon assembly at the outer membrane by using its R domain to bind to a vitamin B-12 receptor BtuB with high affinity, and subsequently searching for OmpF with its disordered N-terminus. Once the colicin locates an available OmpF trimer, its N-terminus would thread through the OmpF lumen to bind to TolB. The binding of colicin E9 to BtuB has been extensively studied: the thermodynamics of the binding are known ($K_d \approx$ 2.5 nM[50]), as is the crystal structure of BtuB bound to the R domain of colicin E3 (identical to that of E9). However, the mechanism of OmpF recruitment and the involvement of the movement of the folded part of the T domain in this step of translocon formation remain unclear. It is thought that the colicin uses its R domain as a fishing rod from BtuB, while its T domain acts as a fishing line to probe the surrounding area for OmpF[48]. Involvement of the movement of the helices of the R-domain relative to each other in this process was debated. While some literature suggests that the R domain becomes more disordered after binding to BtuB[48], more recent literature argues that this is not the case[50]. It is also known, that the T domain undergoes structural rearrangements and likely unfolds upon binding to OmpF and TolB to initiate translocation[74]. As rearrangement of the T domain was prevented by the disulfide bridge between residues 85 and 104, this study was directed at sampling the conformation of colicin E9 bound to both BtuB and OmpF before translocation begins. The results presented in section 4.2.3 show that lysine specific crosslinking is a valid approach to sample structural conformations and protein-protein interactions within colicin translocons. The crosslinks identified between each pair of molecules show that the translocon of colicin E9 is a compact structure where each of the proteins at the outer membrane interacts with

the other two. The interfaces between each pair of proteins could be predicted using the crosslinks and previous data as distance restraints. However, no matter how these proteins were arranged, it was impossible to fit all three interfaces together. The major culprit was the OmpF-BtuB cross-link, which agreed with previous MD simulation data, but if satisfied, would put the T domain of the colicin away from the OmpF trimer. One possible explanation for this would be potential 'ball-and socket' type of rotation of the colicin R domain at the BtuB binding site, although this sort of behavior has not been observed previously. It is possible that the crystal structure represents the initial BtuB binding interface of colicin E9 as it would be the most stable structure, while upon interaction of the disordered N-terminal domain with OmpF and TolB this angle may change, weakening the interaction of the R-domain with BtuB, making subsequent dissociation possible. During this process, the colicin would have to undergo some conformational changes as well. This is supported by the crosslinks that were identified within colicin E9, which suggest that the orientation and the angle of the T domain relative to the R domain are different when the colicin is alone or in complex with BtuB relative to when it is part of the assembled translocon. Moreover, our data suggest that there are multiple ways the colicin can bridge BtuB and OmpF, depending on the orientation of the trimer relative to BtuB, as well as the orientation of BtuB relative to each of the OmpF monomers. It is however unclear whether the OmpF and BtuB interaction occurs before or after the binding of the colicin to both proteins and whether this sort of movement of the R domain relative to BtuB is energetically favourable. Therefore, at this point, this is only one of several hypotheses of colicin E9 translocon formation at the outer membrane, which could be tested in the future through several important experiments described in the next section.

4.3.2 Possible future work

So far, most of the work on the translocon assembly has been done *in vitro* (thermodynamic and kinetic studies), and any structural work captured only assembled parts of the translocon (OmpF-TolB cryo-EM structure, BtuB and the minimum R domain of the colicin). For this study as well, the translocon was already assembled at the outer membrane prior to purification and addition of TolB, so it is unclear what is the role of the TolB binding in translocon assembly and whether the interaction between OmpF and BtuB occurs prior to or during translocon assembly. OmpF and BtuB are sufficient for translocon assembly *in vitro,* while this may be different *in vivo* if these proteins are not associated or close to each other prior to colicin binding Therefore, *in vivo* studies monitoring the translocon assembly could provide valuable insight. However, small timescales, difficulties in imaging single particles using current methods, and translocation of colicins, make the design of such studies challenging

We know that outer membrane proteins are arranged into islands of different shapes and sizes, which are created in the middle of the growing cells, and are pushed towards the poles as the cells grow[141]. However, we do not yet know anything about the composition of these islands. So one of the most interesting and potentially achievable future experiments would be determining the cell surface proteome of *E. coli* and using lysine specific crosslinking to discover any proteins which are located close to each other. This could help us not only to shine a light on association between OmpF and BtuB in the cells prior to colicin binding but also uncover potential relationships between other proteins in the outer membrane. The extracellular surface of most outer membrane proteins is lined with lysine residues; therefore interactions between the β-barrels could be captured with NHS ester crosslinkers of medium length. The main difficulties with these experiments would be enrichment and extraction of cross-linked and labeled proteins from the cells, but they

could be easily addressed. The identification of proteins at the outer membrane of *E. coli* could be achieved by use of cleavable biotin-labeled mono-functional NHS linkers (commonly used for mammalian cell surface proteome analysis) to label the proteins and using sodium deoxycholate to extract them from the membrane prior to analysis by label-free quantitative mass spectrometry.

The crosslinking experiment would be much more complicated as not only would the crosslinked proteins need to be identified, but also the peptides containing the crosslinked lysines. The process of enrichment and identification of the peptides would have to be different to that used in this study as the initial sample would be much more complex (likely at least 10 times as many proteins as involved in the translocon formation), and consequently the search database would be much larger. The simple method of sequencing only highly charged peptides would be insufficient to enrich the cross-linked samples, so a purification step would be necessary. One possible solution to this issue is the use of tri-functional cross-linkers, which have a cleavable biotin tag in addition to NHS-ester containing cross-linkers; although currently available reagents of this type are not soluble in water, and need to be dissolved in 100% dimethyl sulfoxide (DMSO) prior to use. This may be an issue as DMSO may affect membrane permeability and as a result allow crosslinks to form between proteins located in the periplasm as well as the inner membrane further complicating the analysis. However, with currently rapid development of various crosslinkers this issue may soon be overcome and allow these experiments to be done.

As well as a requirement for knowledge of organisation of the outer membrane, the previous section highlighted the need for more studies on the dynamics of the BtuB-ColE9 complex before and after binding to OmpF *in vitro* and *in vivo*. One way of achieving this would be designing FRET fluorophore pairs on colicin E9 R domain and on BtuB loops

using the information provided by the crystal structure and monitoring the efficiency of the energy transfer during translocon assembly. Due to development of new specific labeling techniques potentially applicable to cell-surface proteins[192], these experiments could be conducted *in vitro* and *in vivo.* A negative change in the FRET efficiency would support the hypothesis described above. In addition to FRET pairs between colicin E9 and BtuB, it may be useful to monitor the distance between the T domain and the top of the R domain of the colicin E9 near the hinge region, which separates the two domains *in vivo*. OmpF binding may also be monitored by FRET. The design of positioning of the fluorophore pairs could be informed by the model of ColE9-OmpF interface. Mutations in the TolB box of the colicin (or TolB knockouts) for the *in vivo* experiments could also help understand its involvement in formation of the colicin E9 translocon.

5. Major conclusions and impact of these studies on further research

The translocation of colicins across the outer membrane is a remarkable phenomenon, which could be exploited for effective species-specific drug delivery and design. However, before the design of such agents is possible, we need to have a comprehensive knowledge and understanding of the mechanism of translocation, of regions and structures which colicins require for cytotoxicity depending on species they target and cellular machineries they use, and of potential routes through which the target cells may become resistant. The focus of this study was testing whether it is possible to use colicins as delivery vehicles for toxic peptides and investigating the topology of the colicin E9 translocon. The major outcomes of this work are summarized below.

5.1 Using colicins to deliver toxic peptides into the periplasm

In this study I've used the disordered N-terminus of colicin N to deliver a toxic BamD binding peptide into the *E. coli* B^E3000 cells. I found that the insertion of the eight amino acid peptide into the N-terminus between the OmpF binding site and the TolA-box reduces the translocation efficiency of the constructs, while inclusion of the peptide in the region between the TolA-box and the folded domain had no effect on translocation efficiency. Therefore the 20-amino acid region after the TolA-box is a good position for insertion of transported toxic peptides.

Inclusion of the BamD peptide improves the cytotoxicity of the colicin N-TR construct, however it does not have a significant effect on its minimum inhibitory concentration, as it is likely to require many translocated molecules per cell to have an effect. The data

presented in chapter 3 show that it is possible to deliver cytotoxic peptides using colicin N-terminal domain, however further experiments are required to understand the effect produced by the BamD binding peptide. The results of the study also show that there are differences in colicin activity depending on the media in which the cells are grown. Further investigations into efficacy of colicins in different environments may aid discovery of new environment specific targets for antibiotics.

5.2 Mapping the topology of the translocon

Translocon formation is the first step of the translocation of colicins, therefore knowing the mechanism of translocon assembly and its structure is key to understanding the process of colicin translocation, and potential opportunities for future drug design. The results described in chapter 4 demonstrate that lysine-specific crosslinking coupled with mass spectrometry can be applied to gain insight into structural organization of the translocon, which is difficult to study with conventional structural methods such as x-ray crystallography.

The crosslinking experiments have shown that the colicin E9 translocon is a compact entity with three inter-protein interfaces: one between the T domain of the colicin and OmpF, one between OmpF and BtuB and one between the R domain of the colicin and BtuB. While the crystal structure of the interface between BtuB and the colicin is known, and two interfaces of the BtuB-OmpF interface have been previously predicted, the interface between the T domain of the colicin E9 and OmpF was modeled using identified crosslinks as distance restraints for docking. These interfaces cannot be satisfied simultaneously, as the BtuB-OmpF interface places the colicin T domain away from the OmpF trimer implying a rotation of the colicin R domain when it is bound to BtuB. This is unlikely to occur

naturally and may be caused by OmpF binding. The intra-molecular crosslinks within colicin E9 identified in this study suggest that the structure of the colicin is slightly different in solution (and in the crystal) compared to when it is part of the translocon. However, this conformational change cannot be modeled at present, as it appears that it is not the same in every translocon.

Based on our data, we present several possible models of colicin E9 translocons and suggest some changes to the current hypothesis for translocon assembly, which should be considered in the future research. The main proposed change is introduction of rotational movement of colicin E9 R domain relative to BtuB, which may aid dissociation of the R domain from BtuB during its translocation. Further *in vitro* and *in vivo* cross-linking experiments as well as *in vivo* FRET studies could be conducted to test this hypothesis, however these crosslinking data would be most useful in combination with structures (even at low resolutions) obtained through cryo-EM.

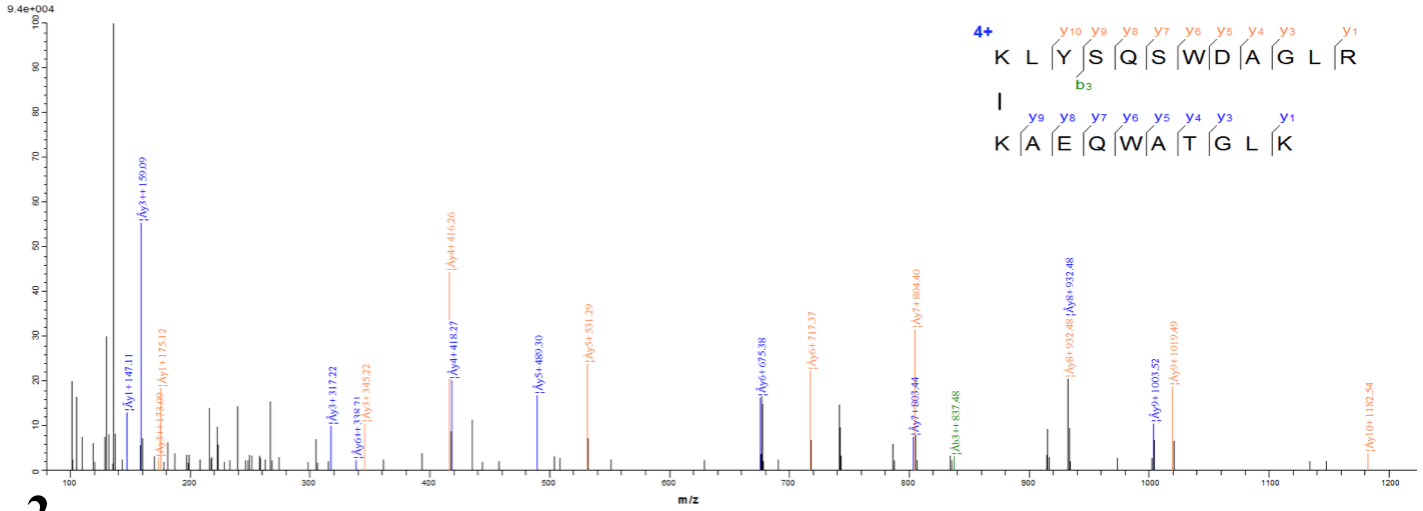
Word count 23,566

Appendix

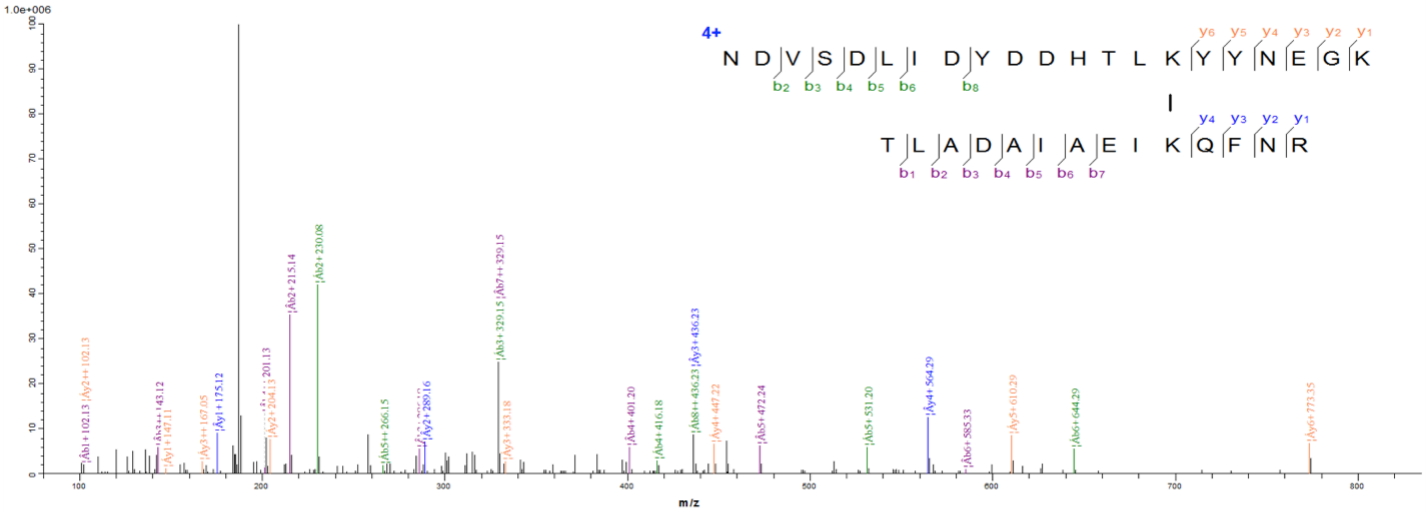
Figure I: Best identified spectra for inter-protein crosslinks numbered from 1-8 as referenced in the text (see table below). These spectra were drawn using pLink plugin – pLabel. The sequences of identified peptides along with the charge state of the parent (precursor ion) and y and b ions, which were matched to the theoretical spectra to produce the identification, are shown in the top right corner. The x-axis of each spectrum contains information on m/z of a fragment ion, while the y-axis contains information on relative intensity of each fragment ion in the sample. Highest intensity absolute value for each of the spectra is shown at the top of the y-axis. Oxidised methionine residue is highlighted in red

Appendix Ref. No	Linked residues		E-values
1	BtuB (K264)	OmpF (K232)	2.95E-14
2	ColE9 (K374)	BtuB (K472)	2.41E-10
3	ColE9 (K147)	OmpF (K231)	4.56E-04
4	ColE9 (K177)	OmpF (K231)	1.78E-06
5	ColE9 (K277)	OmpF (K47)	1.32E-09
6	ColE9 (K347)	OmpF (K231)	2.60E-04*
7	ColE9 (S2)	TolB (K117)	1.03E-09
8	ColE9 (S2)	TolB (243)	7.07E-06

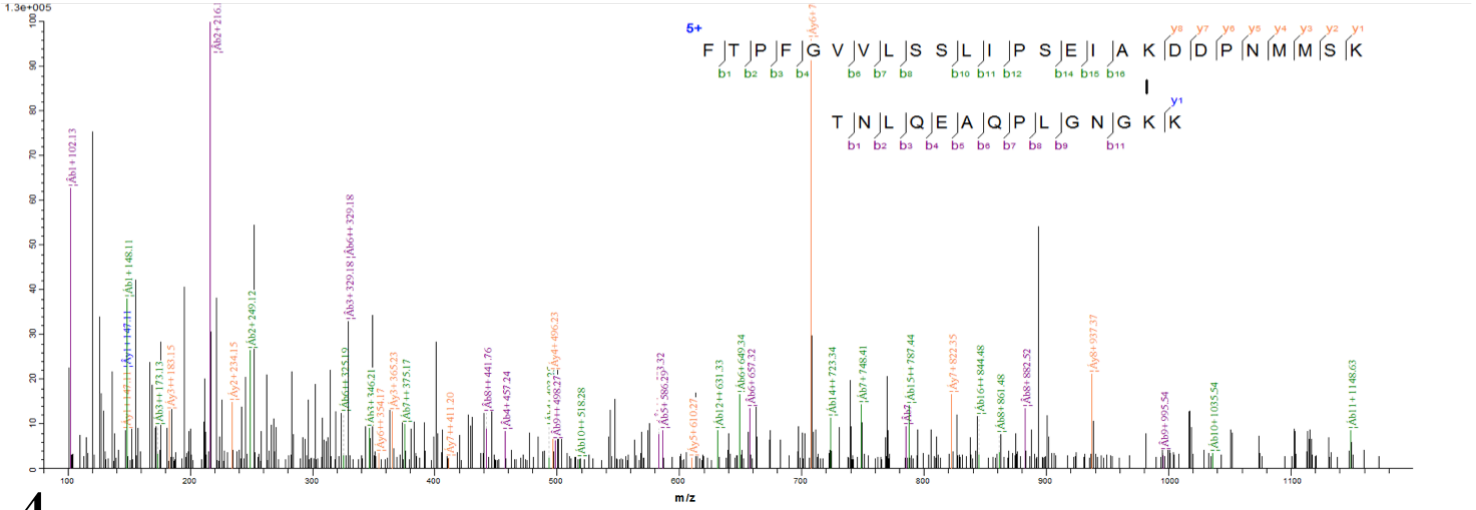
1


9.4e+004
4+
K L Y S Q S W D A G L R
b3
I
K A E Q W A T G L K
m/z

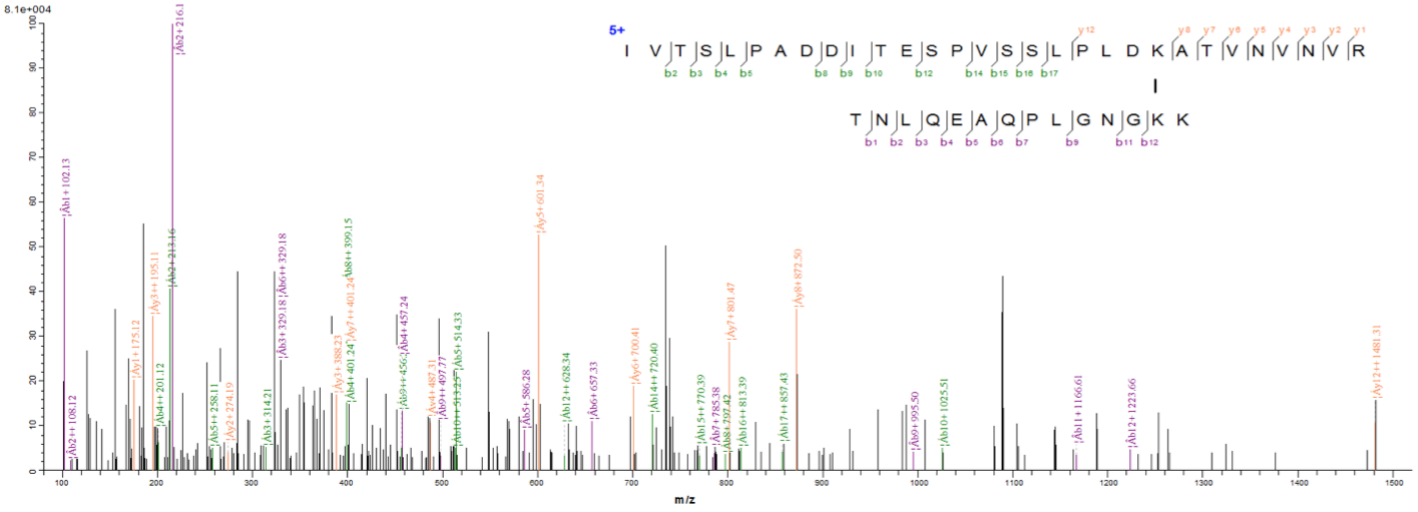
2


1.0e+006
4+
N D V S D L I D Y D D H T L K Y Y N E G K
b2 b3 b4 b5 b6 b8
I
T L A D A I A E I K Q F N R
b1 b2 b3 b4 b5 b6 b7
m/z

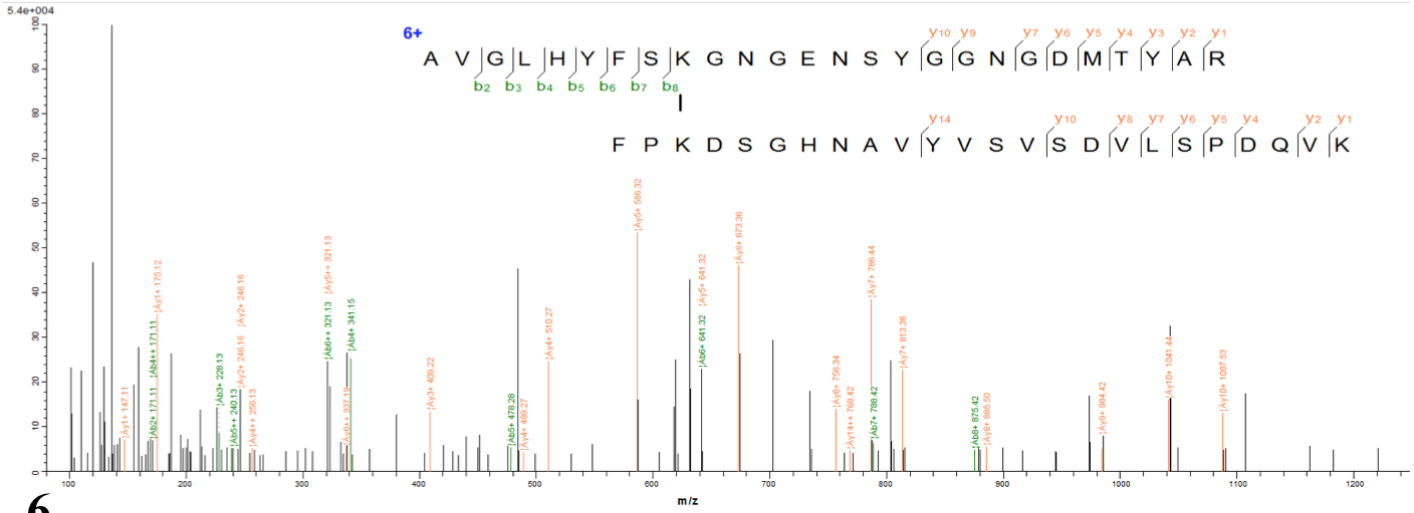
3


1.3e+005
5+
F T P F G V V L S S L I P S E I A K D D P N M M S K
T N L Q E A Q P L G N G K K
m/z

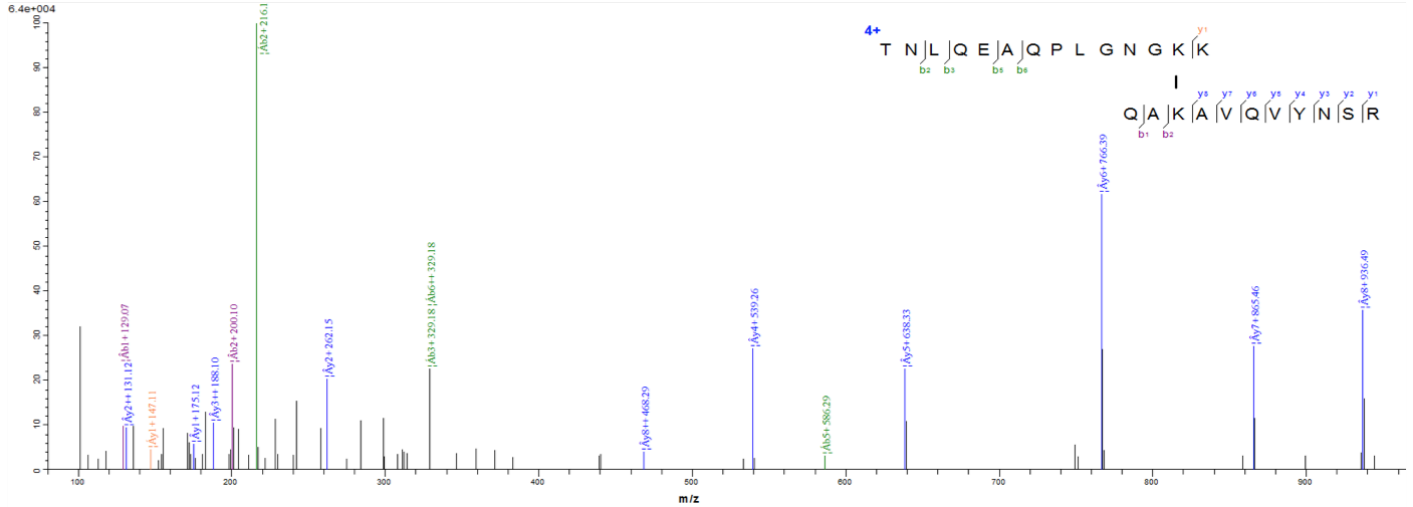
4


8.1e+004
5+
I V T S L P A D D I T E S P V S S L P L D K A T V N V N V R
T N L Q E A Q P L G N G K K
m/z

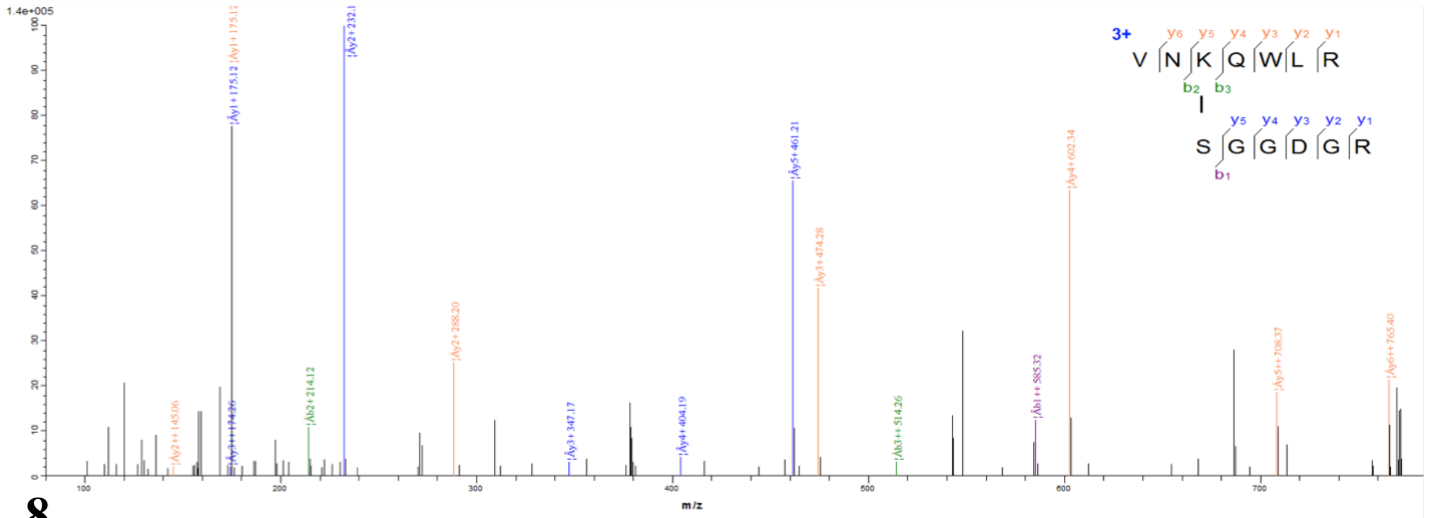
5


6+
A V G L H Y F S K G N G E N S Y G G N G D M T Y A R
F P K D S G H N A V Y V S V S D V L S P D Q V K
m/z

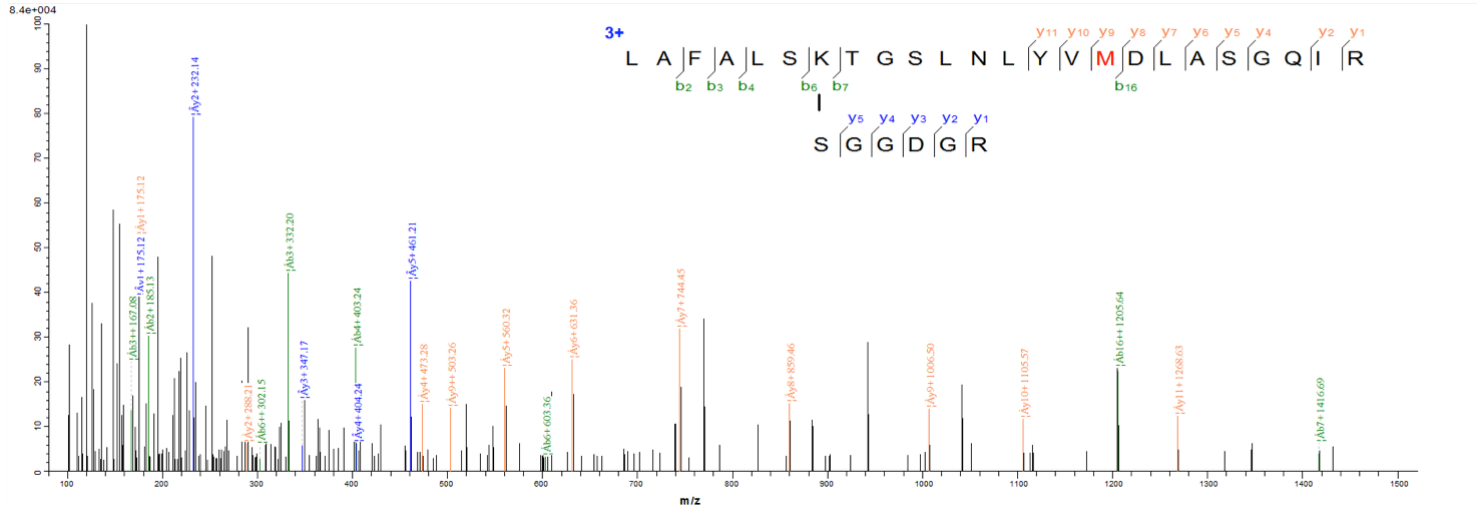
6


4+
T N L Q E A Q P L G N G K K
Q A K A V Q V Y N S R
m/z

7


1.4e+005
3+
V N K Q W L R
S G G D G R
m/z

8


8.4e+004
3+
L A F A L S K T G S L N L Y V M D L A S G Q I R
S G G D G R
m/z

References

1. A, G. Sur un remarquable example d'antagonisme entre deux souches de colibacille. *Compt. Rend. Soc. Biol.* **93,** 1040–2 (1925).
2. Fredericq, P. & Gratia, A. Diversité des souches antibiotiques de B. coli et etendue variable de leur champ d'action. *C.R. Soc. Biol.* 1032-1033 (1946).
3. Davies, J. K. & Reeves, P. Genetics of resistance to colicins in Escherichia coli K-12: cross-resistance among colicins of group B. *J. Bacteriol.* **123,** 96–101 (1975).
4. Davies, J. K. & Reeves, P. Genetics of resistance to colicins in Escherichia coli K-12: cross-resistance among colicins of group A. *J. Bacteriol.* **123,** 102–17 (1975).
5. Cascales, E. *et al.* Colicin biology. *Microbiol. Mol. Biol. Rev.* **71,** 158–229 (2007).
6. Jakes, K. S. & Cramer, W. A. Border Crossings: Colicins and Transporters. *Annu. Rev. Genet.* **46,** 209–231 (2012).
7. Šmajs, D. *et al.* Bacteriocin synthesis in uropathogenic and commensal Escherichia coli: colicin E1 is a potential virulence factor. *BMC Microbiol.* **10,** 288 (2010).
8. Šmarda, J. & Obdržálek, V. Incidence of colicinogenic strains among humanEscherichia coli. *J. Basic Microbiol.* **41,** 367–374 (2001).
9. De Witt, W. & Helinski, D. R. Colicinogenic factor from a non-induced and a mitomycin C-induced Proteus strain. *J. Mol. Biol.* **13,** 692–703 (1965).
10. Hardy, K. G., Meynell, G. G., Dowman, J. E. & Spratt, B. G. Two major groups of colicin factors: their evolutionary significance. *Mol. Gen. Genet.* **125,** 217–30 (1973).
11. Isaacson, R. E. & Konisky, J. Studies on the regulation of colicin Ib synthesis: replication of the Col Ib-P9 plasmid during colicin induction. *Antimicrob. Agents Chemother.* **6,** 848–52 (1974).
12. Durkacz, B. W., Kennedy, C. K. & Sherratt, D. J. Plasmid replication and the induced synthesis of colicins E1 and E2 in Escherichia coli. *J. Bacteriol.* **117,** 940–6 (1974).
13. Riley, M. A. Positive selection for colicin diversity in bacteria. *Mol. Biol. Evol.* **10,** 1048–59 (1993).
14. Riley, M. A. Molecular mechanisms of colicin evolution. *Mol. Biol. Evol.* **10,** 1380–95 (1993).
15. Pugsley, A. P. The immunity and lysis genes of ColN plasmid pCHAP4. *MGG Mol. Gen. Genet.* **211,** 335–341 (1988).
16. Goldman, K., Suit, J. L. & Kayalar, C. *Identification of the plasmid-encoded immunity protein for colicin El in the inner membrane of Escherichia coli*. *FEBS Letters* **190,** (No longer published by Elsevier, 1985).
17. Lloubes, R. P., Chartier, M. J., Journet, A. M., Varenne, S. G. & LAZDUNSKI, C. J. Nucleotide sequence of the gene for the immunity protein to colicin A. Analysis of codon usage of immunity proteins as compared to colicins. *Eur. J. Biochem.* **144,** 73–78 (1984).
18. Olschläger, T. & Braun, V. Sequence, expression, and localization of the immunity protein for colicin M. *J. Bacteriol.* **169,** 4765–9 (1987).
19. Lloubes, R., Baty, D. & Lazdunski, C. The promoters of the genes for colicin production, release and immunity in the ColA plasmid: effects of convergent transcription and Lex A protein. *Nucleic Acids Res.* **14,** 2621–36 (1986).
20. Alderman, E. M., Dills, S. S., Melton, T. & Dobrogosz, W. J. Cyclic adenosine 3',5'-monophosphate regulation of the bacteriophage T6/colicin K receptor in

Escherichia coli. *J. Bacteriol.* **140,** 369–76 (1979).
21. Salles, B. & Weinstock, G. M. Interaction of the CRP-cAMP complex with the cea regulatory region. *Mol. Gen. Genet.* **215,** 537–42 (1989).
22. Butala, M. *et al.* Double locking of an Escherichia coli promoter by two repressors prevents premature colicin expression and cell lysis. *Mol. Microbiol.* **86,** 129–39 (2012).
23. Kamenšek, S. *et al.* Silencing of DNase Colicin E8 Gene Expression by a Complex Nucleoprotein Assembly Ensures Timely Colicin Induction. *PLoS Genet.* **11,** e1005354 (2015).
24. Hol, F. J. H., Voges, M. J., Dekker, C. & Keymer, J. E. Nutrient-responsive regulation determines biodiversity in a colicin-mediated bacterial community. *BMC Biol.* **12,** 68 (2014).
25. Konisky, J. & Richards, F. M. Characterization of colicin Ia and colicin Ib. Purification and some physical properties. *J. Biol. Chem.* **245,** 2972–8 (1970).
26. Papadakos, G., Wojdyla, J. A. & Kleanthous, C. Nuclease colicins and their immunity proteins. *Q. Rev. Biophys.* **45,** 57–103 (2012).
27. Schramm, E., Mende, J., Braun, V. & Kamp, R. M. Nucleotide sequence of the colicin B activity gene cba: consensus pentapeptide among TonB-dependent colicins and receptors. *J. Bacteriol.* **169,** 3350–7 (1987).
28. Schwartz, S. A. & Helinski, D. R. Purification and characterization of colicin E1. *J. Biol. Chem.* **246,** 6318–27 (1971).
29. Herschman, H. R. & Helinski, D. R. Comparative study of the events associated with colicin induction. *J. Bacteriol.* **94,** 691–9 (1967).
30. Herschman, H. R. & Helinski, D. R. Purification and characterization of colicin E2 and colicin E3. *J. Biol. Chem.* **242,** 5360–8 (1967).
31. Timmis, K. Purification and characterization of colicin D. *J. Bacteriol.* **109,** 12–20 (1972).
32. Hands, S. L. *et al.* Interactions of TolB with the translocation domain of colicin E9 require an extended TolB box. *J. Bacteriol.* **187,** 6733–41 (2005).
33. Housden, N. G. *et al.* Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science* **340,** 1570–4 (2013).
34. Ridley, H. & Lakey, J. H. Antibacterial toxin colicin N and phage protein G3p compete with TolB for a binding site on TolA. *Microbiology* **161,** 503–15 (2015).
35. Garinot-Schneider, C., Penfold, C. N., Moore, G. R., Kleanthous, C. & James, R. Identification of residues in the putative TolA box which are essential for the toxicity of the endonuclease toxin colicin E9. *Microbiology* 2931–8 (1997). doi:10.1099/00221287-143-9-2931
36. Mende, J. & Braun, V. Import-defective colicin B derivatives mutated in the TonB box. *Mol. Microbiol.* **4,** 1523–1533 (1990).
37. Jakes, K. S. & Finkelstein, A. The colicin Ia receptor, Cir, is also the translocator for colicin Ia. *Mol. Microbiol.* **75,** 567–78 (2010).
38. Klein, A. *et al.* Structural and biophysical analysis of nuclease protein antibiotics. *Biochem. J.* **473,** 2799–2812 (2016).
39. Soelaiman, S., Jakes, K., Wu, N., Li, C. & Shoham, M. Crystal structure of colicin E3: implications for cell entry and ribosome inactivation. *Mol.Cell* **8,** 1053–1062 (2001).
40. Wiener, M., Freymann, D., Ghosh, P. & Stroud, R. M. Crystal structure of colicin Ia. *Nature* **385,** 461–4 (1997).
41. Yajima, S. *et al.* Structural basis for sequence-dependent recognition of colicin E5 tRNase by mimicking the mRNA-tRNA interaction. *Nucleic Acids Res.* **34,** 6074–82

(2006).
42. Graille, M., Mora, L., Buckingham, R. H., van Tilbeurgh, H. & de Zamaroczy, M. Structural inhibition of the colicin D tRNase by the tRNA-mimicking immunity protein. *EMBO J.* **23,** 1474–82 (2004).
43. Gérard, F. *et al.* X-ray structure and site-directed mutagenesis analysis of the Escherichia coli colicin M immunity protein. *J. Bacteriol.* **193,** 205–14 (2011).
44. Kühlmann, U. C., Moore, G. R., James, R., Kleanthous, C. & Hemmings, A. M. Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett.* **463,** 1–2 (1999).
45. Kleanthous, C. *et al.* Structural and mechanistic basis of immunity toward endonuclease colicins. *Nat. Struct. Biol.* **6,** 243–52 (1999).
46. Shub, D. A., Goodrich-Blair, H. & Eddy, S. R. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem. Sci.* **19,** 402–4 (1994).
47. Keeble, A. H., Maté, M. J. & Kleanthous, C. in *Homing Endonucleases and Inteins* 49–65 (Springer-Verlag, 2005). doi:10.1007/3-540-29474-0_4
48. Kurisu, G. *et al.* The structure of BtuB with bound colicin E3 R-domain implies a translocon. *Nat. Struct. Biol.* **10,** 948–54 (2003).
49. Pommer, A. J. *et al.* Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J. Mol. Biol.* **314,** 735–49 (2001).
50. Housden, N. G., Loftus, S. R., Moore, G. R., James, R. & Kleanthous, C. Cell entry mechanism of enzymatic bacterial colicins: porin recruitment and the thermodynamics of receptor binding. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 13849–54 (2005).
51. Carr, S., Penfold, C. N., Bamford, V., James, R. & Hemmings, A. M. The structure of TolB, an essential component of the tol-dependent translocation system, and its protein–protein interaction with the translocation domain of colicin E9. *Structure* **8,** 57–66 (2000).
52. Wallis, R., Moore, G. R., James, R. & Kleanthous, C. Protein-Protein Interactions in Colicin E9 DNase-Immunity Protein Complexes. 1. Diffusion-Controlled Association and Femtomolar Binding for the Cognate Complex. *Biochemistry* **34,** 13743–13750 (1995).
53. Vankemmelbeke, M. *et al.* Energy-dependent immunity protein release during tol-dependent nuclease colicin translocation. *J. Biol. Chem.* **284,** 18932–41 (2009).
54. Bourdineaud, J. P., Fierobe, H. P., Lazdunski, C. & Pagès, J. M. Involvement of OmpF during reception and translocation steps of colicin N entry. *Mol. Microbiol.* **4,** 1737–43 (1990).
55. Vetter, I. R. *et al.* Crystal structure of a colicin N fragment suggests a model for toxicity. *Structure* **6,** 863–74 (1998).
56. Johnson, C. L., Ridley, H., Pengelly, R. J., Salleh, M. Z. & Lakey, J. H. The unstructured domain of colicin N kills Escherichia coli. *Mol. Microbiol.* **89,** 84–95 (2013).
57. Raggett, E. M., Bainbridge, G., Evans, L. J., Cooper, A. & Lakey, J. H. Discovery of critical Tol A-binding residues in the bactericidal toxin colicin N: a biophysical approach. *Mol. Microbiol.* **28,** 1335–43 (1998).
58. Johnson, C. L. *et al.* The antibacterial toxin colicin N binds to the inner core of lipopolysaccharide and close to its translocator protein. *Mol. Microbiol.* **92,** 440–52 (2014).
59. Mock, M. & Pugsley, A. P. The BtuB group col plasmids and homology between the colicins they encode. *J. Bacteriol.* **150,** 1069–76 (1982).

60. Luan, B., Carr, R., Caffrey, M. & Aksimentiev, A. The effect of calcium on the conformation of cobalamin transporter BtuB. *Proteins Struct. Funct. Bioinforma.* **78,** 1153–1162 (2010).
61. Gumbart, J., Wiener, M. C. & Tajkhorshid, E. Coupling of calcium and substrate binding through loop alignment in the outer-membrane transporter BtuB. *J. Mol. Biol.* **393,** 1129–42 (2009).
62. Zakharov, S. D. & Cramer, W. A. On the mechanism and pathway of colicin import across the E. Coli outer membrane. *Front. Biosci.* **9,** 1311–7 (2004).
63. Buchanan, S. K. *et al.* Structure of colicin I receptor bound to the R-domain of colicin Ia: implications for protein import. *EMBO J.* **26,** 2594–604 (2007).
64. Fuller-Schaefer, C. A. & Kadner, R. J. Multiple extracellular loops contribute to substrate binding and transport by the Escherichia coli cobalamin transporter BtuB. *J. Bacteriol.* **187,** 1732–9 (2005).
65. Sharma, O. *et al.* Structure of the complex of the colicin E2 R-domain and its BtuB receptor. The outer membrane colicin translocon. *J. Biol. Chem.* **282,** 23163–70 (2007).
66. Chimento, D. P., Mohanty, A. K., Kadner, R. J. & Wiener, M. C. Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. *Nat. Struct. Biol.* **10,** 394–401 (2003).
67. Payne, M. A. *et al.* Biphasic binding kinetics between FepA and its ligands. *J. Biol. Chem.* **272,** 21950–5 (1997).
68. Nikaido, H. Antibiotic Resistance Caused by Gram-Negative Multidrug Efflux Pumps. *Clin. Infect. Dis.* **27,** S32–S41 (1998).
69. Tikhonova, E. B. & Zgurskaya, H. I. AcrA, AcrB, and TolC of Escherichia coli form a stable intermembrane multidrug efflux complex. *J. Biol. Chem.* **279,** 32116–32124 (2004).
70. Zakharov, S. D., Wang, X. S. & Cramer, W. A. The Colicin E1 TolC-Binding Conformer: Pillar or Pore Function of TolC in Colicin Import? doi:10.1021/acs.biochem.6b00621
71. Baboolal, T. G. *et al.* Colicin N binds to the periphery of its receptor and translocator, outer membrane protein F. *Structure* **16,** 371–9 (2008).
72. Bainbridge, G., Armstrong, G. A., Dover, L. G., Whelan, K. F. & Lakey, J. H. Displacement of OmpF loop 3 is not required for the membrane translocation of colicins N and A in vivo. *FEBS Lett.* **432,** 117–22 (1998).
73. Zhang, Y. *et al.* Investigating early events in receptor binding and translocation of colicin E9 using synchronized cell killing and proteolytic cleavage. *J. Bacteriol.* **190,** 4342–50 (2008).
74. Vankemmelbeke, M., Housden, N. G., James, R., Kleanthous, C. & Penfold, C. N. Immunity protein release from a cell-bound nuclease colicin complex requires global conformational rearrangement. *Microbiologyopen* **2,** 853–61 (2013).
75. Lloubès, R. *et al.* The Tol-Pal proteins of the Escherichia coli cell envelope: an energized system required for outer membrane integrity? *Res. Microbiol.* **152,** 523–529 (2001).
76. Gerding, M. A., Ogata, Y., Pecora, N. D., Niki, H. & de Boer, P. A. J. The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in E. coli. *Mol. Microbiol.* **63,** 1008–25 (2007).
77. Cascales, E., Lloubès, R. & Sturgis, J. N. The TolQ-TolR proteins energize TolA and share homologies with the flagellar motor proteins MotA-MotB. *Mol. Microbiol.* **42,** 795–807 (2001).

78. Walburger, A., Lazdunski, C. & Corda, Y. The Tol/Pal system function requires an interaction between the C-terminal domain of TolA and the N-terminal domain of TolB. *Mol. Microbiol.* **44,** 695–708 (2002).
79. Schendel, S. L., Click, E. M., Webster, R. E. & Cramer, W. A. The TolA Protein Interacts with Colicin E1 Differently than with Other Group A Colicins. *J. Bacteriol.* **179,** 3683–3690 (1997).
80. Penfold, C. N., Li, C., Zhang, Y., Vankemmelbeke, M. & James, R. Colicin A binds to a novel binding site of TolA in the Escherichia coli periplasm. *Biochem. Soc. Trans.* **40,** 1469–74 (2012).
81. Braun, V. Energy-coupled transport and signal transduction through the Gram-negative outer membrane via TonB-ExbB-ExbD-dependent receptor proteins. *FEMS Microbiol. Rev.* **16,** 295–307 (1995).
82. Braun, V., G□nter, K. & Hantke, K. Transport of iron across the outer membrane. *Biol. Met.* **4,** 14–22 (1991).
83. Braun, V. The structurally related exbB and tolQ genes are interchangeable in conferring tonB-dependent colicin, bacteriophage, and albomycin sensitivity. *J. Bacteriol.* **171,** 6387–90 (1989).
84. Bouveret, E., Rigal, A., Lazdunski, C. & Bénédetti, H. The N-terminal domain of colicin E3 interacts with TolB which is involved in the colicin translocation step. *Mol. Microbiol.* **23,** 909–20 (1997).
85. Bradbeer, C. The proton motive force drives the outer membrane transport of cobalamin in Escherichia coli. *J. Bacteriol.* **175,** 3146–50 (1993).
86. Evans, J. S., Levine, B. A., Trayer, I. P., Dorman, C. J. & Higgins, C. F. Sequence-imposed structural constraints in the TonB protein of E. coli. *FEBS Lett.* **208,** 211–6 (1986).
87. Moeck, G. S. & Letellier, L. Characterization of in vitro interactions between a truncated TonB protein from Escherichia coli and the outer membrane receptors FhuA and FepA. *J. Bacteriol.* **183,** 2755–64 (2001).
88. Heller, K. J., Kadner, R. J. & Günther, K. Suppression of the btuB451 mutation by mutations in the tonB gene suggests a direct interaction between TonB and TonB-dependent receptor proteins in the outer membrane of Escherichia coli. *Gene* **64,** 147–53 (1988).
89. Günter, K. & Braun, V. In vivo evidence for FhuA outer membrane receptor interaction with the TonB inner membrane protein of Escherichia coli. *FEBS Lett.* **274,** 85–8 (1990).
90. Tuckman, M. & Osburne, M. S. In vivo inhibition of TonB-dependent processes by a TonB box consensus pentapeptide. *J. Bacteriol.* **174,** 320–3 (1992).
91. Mende, J. & Braun, V. Import-defective colicin B derivatives mutated in the TonB box. *Mol. Microbiol.* **4,** 1523–33 (1990).
92. Chauleau, M., Mora, L., Serba, J. & de Zamaroczy, M. FtsH-dependent processing of RNase colicins D and E3 means that only the cytotoxic domains are imported into the cytoplasm. *J. Biol. Chem.* **286,** 29397–407 (2011).
93. Cavard, D. & Lazdunski, C. Colicin cleavage by OmpT protease during both entry into and release from Escherichia coli cells. *J. Bacteriol.* **172,** 648–52 (1990).
94. LAKEY, J. H. *et al.* Membrane insertion of the pore-forming domain of colicin A. A spectroscopic study. *Eur. J. Biochem.* **196,** 599–607 (1991).
95. Ridley, H., Johnson, C. L. & Lakey, J. H. Interfacial interactions of pore-forming colicins. *Adv. Exp. Med. Biol.* **677,** 81–90 (2010).
96. Kim, Y., Valentine, K., Opella, S. J., Schendel, S. L. & Cramer, W. A. Solid-state NMR studies of the membrane-bound closed state of the colicin E1 channel domain

in lipid bilayers. *Protein Sci.* **7,** 342–8 (1998).
97. Ho, D. *et al.* Membrane topology of the colicin E1 channel using genetically encoded fluorescence. *Biochemistry* **50,** 4830–42 (2011).
98. Zakharov, S. D. *et al.* Membrane-bound state of the colicin E1 channel domain as an extended two-dimensional helical array. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 4282–7 (1998).
99. Pulagam, L. P. & Steinhoff, H.-J. Acidic pH-induced membrane insertion of colicin A into E. coli natural lipids probed by site-directed spin labeling. *J. Mol. Biol.* **425,** 1782–94 (2013).
100. Greig, S. L., Radjainia, M. & Mitra, A. K. Oligomeric structure of colicin ia channel in lipid bilayer membranes. *J. Biol. Chem.* **284,** 16126–34 (2009).
101. Dunkel, S. *et al.* In vivo EPR on spin labeled colicin A reveals an oligomeric assembly of the pore-forming domain in E. coli membranes. *Phys. Chem. Chem. Phys.* **17,** 4875–4878 (2015).
102. Bullock, J. O., Kolen, E. R. & Shear, J. L. Ion selectivity of colicin E1: II. Permeability to organic cations. *J. Membr. Biol.* **128,** 1–16 (1992).
103. Bullock, J. O. & Kolen, E. R. Ion selectivity of colicin E1: III. Anion permeability. *J. Membr. Biol.* **144,** 131–45 (1995).
104. Nardi, A., Slatin, S. L., Baty, D. & Duché, D. The C-terminal half of the colicin A pore-forming domain is active in vivo and in vitro. *J. Mol. Biol.* **307,** 1293–303 (2001).
105. Qiu, X. Q., Jakes, K. S., Kienker, P. K., Finkelstein, A. & Slatin, S. L. Major transmembrane movement associated with colicin Ia channel gating. *J. Gen. Physiol.* **107,** 313–28 (1996).
106. Slatin, S. L., Duché, D., Kienker, P. K. & Baty, D. Gating movements of colicin A and colicin Ia are different. *J. Membr. Biol.* **202,** 73–83 (2004).
107. Peraro, M. D. & van der Goot, F. G. Pore-forming toxins: ancient, but never really out of fashion. *Nat. Rev. Microbiol.* **14,** 77–92 (2015).
108. Nardi, A., Corda, Y., Baty, D. & Duché, D. Colicin A immunity protein interacts with the hydrophobic helical hairpin of the colicin A channel domain in the Escherichia coli inner membrane. *J. Bacteriol.* **183,** 6721–5 (2001).
109. Pilsl, H. & Braun, V. Evidence that the immunity protein inactivates colicin 5 immediately prior to the formation of the transmembrane channel. *J. Bacteriol.* **177,** 6966–72 (1995).
110. Mate, M. J. & Kleanthous, C. Structure-based Analysis of the Metal-dependent Mechanism of H-N-H Endonucleases. *J. Biol. Chem.* **279,** 34763–34769 (2004).
111. Bowman, C. M., Dahlberg, J. E., Ikemura, T., Konisky, J. & Nomura, M. Specific inactivation of 16S ribosomal RNA induced by colicin E3 in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **68,** 964–8 (1971).
112. Ng, C. L. *et al.* Structural basis for 16S ribosomal RNA cleavage by the cytotoxic domain of colicin E3. *Nat. Struct. Mol. Biol.* **17,** 1241–6 (2010).
113. Ogawa, T. *et al.* A cytotoxic ribonuclease targeting specific transfer RNA anticodons. *Science* **283,** 2097–100 (1999).
114. Tomita, K., Ogawa, T., Uozumi, T., Watanabe, K. & Masaki, H. A cytotoxic ribonuclease which specifically cleaves four isoaccepting arginine tRNAs at their anticodon loops. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 8278–83 (2000).
115. Yajima, S. *et al.* Relation between tRNase activity and the structure of colicin D according to X-ray crystallography. *Biochem. Biophys. Res. Commun.* **322,** 966–73 (2004).
116. Lin, Y.-L., Elias, Y. & Huang, R. H. Structural and mutational studies of the

catalytic domain of colicin E5: a tRNA-specific ribonuclease. *Biochemistry* **44,** 10494–500 (2005).
117. El Ghachi, M. *et al.* Colicin M Exerts Its Bacteriolytic Effect via Enzymatic Degradation of Undecaprenyl Phosphate-linked Peptidoglycan Precursors. *J. Biol. Chem.* **281,** 22761–22772 (2006).
118. Tou, T. *et al.* Colicin M, a peptidoglycan lipid-II-degrading enzyme: potential use for antibacterial means? *Biochem. Soc. Trans.* **40,** (2012).
119. Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochim. Biophys. Acta* **1794,** 808–16 (2009).
120. Li, X.-Z., Plésiat, P. & Nikaido, H. The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clin. Microbiol. Rev.* **28,** 337–418 (2015).
121. Jean, S.-S. *et al.* Carbapenemase-producing Gram-negative bacteria: current epidemics, antimicrobial susceptibility and treatment options. *Future Microbiol.* **10,** 407–25 (2015).
122. Taylor, P. K., Yeung, A. T. Y. & Hancock, R. E. W. Antibiotic resistance in Pseudomonas aeruginosa biofilms: towards the development of novel anti-biofilm therapies. *J. Biotechnol.* **191,** 121–30 (2014).
123. Zacharof, M. P. & Lovitt, R. W. Bacteriocins Produced by Lactic Acid Bacteria A Review Article. *APCBEE Procedia* **2,** 50–56 (2012).
124. Egan, K. *et al.* Bacteriocins: Novel Solutions to Age Old Spore-Related Problems? *Front. Microbiol.* **7,** 461 (2016).
125. Cotter, P. D., Ross, R. P. & Hill, C. Bacteriocins — a viable alternative to antibiotics? *Nat. Rev. Microbiol.* **11,** 95–105 (2012).
126. Bastos, M. do C. de F., Coelho, M. L. V. & Santos, O. C. da S. Resistance to bacteriocins produced by Gram-positive bacteria. *Microbiology* **161,** 683–700 (2015).
127. Budič, M., Rijavec, M., Petkovšek, Z. & Zgur-Bertok, D. Escherichia coli bacteriocins: antimicrobial efficacy and prevalence among isolates from patients with bacteraemia. *PLoS One* **6,** e28769 (2011).
128. Rendueles, O., Beloin, C., Latour-Lambert, P. & Ghigo, J.-M. A new biofilm-associated colicin with increased efficiency against biofilm bacteria. *ISME J.* **8,** 1275–88 (2014).
129. Trautner, B. W., Hull, R. A. & Darouiche, R. O. Colicins prevent colonization of urinary catheters. *J. Antimicrob. Chemother.* **56,** 413–5 (2005).
130. Cutler, S. A., Lonergan, S. M., Cornick, N., Johnson, A. K. & Stahl, C. H. Dietary inclusion of colicin e1 is effective in preventing postweaning diarrhea caused by F18-positive Escherichia coli in pigs. *Antimicrob. Agents Chemother.* **51,** 3830–5 (2007).
131. Tran, E. N. H., Papadopoulos, M. & Morona, R. Relationship between O-antigen chain length and resistance to colicin E2 in Shigella flexneri. *Microbiology* **160,** 589–601 (2014).
132. Clifton, L. A. *et al.* The Effect of Lipopolysaccharide Core Oligosaccharide Size on the Electrostatic Binding of Antimicrobial Proteins to Models of the Gram Negative Bacterial Outer Membrane. *Langmuir* **32,** 3485–94 (2016).
133. Brambilla, L., Morán-Barrio, J. & Viale, A. M. Expression of the Escherichia coli ompW colicin S4 receptor gene is regulated by temperature and modulated by the H-NS and StpA nucleoid-associated proteins. *FEMS Microbiol. Lett.* **352,** 238–44 (2014).
134. Jeanteur, D. *et al.* Structural and functional alterations of a colicin-resistant mutant of OmpF porin from Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10675–9

(1994).
135. Feldgarden, M. & Riley, M. A. The Phenotypic and Fitness Effects of Colicin Resistance in Escherichia coli K-12. *Evolution (N. Y).* **53,** 1019 (1999).
136. Feldgarden, M. & Riley, M. A. High Levels of Colicin Resistance in Escherichia coli. *Evolution (N. Y).* **52,** 1270 (1998).
137. Kleanthous, C., Hemmings, A. M., Moore, G. R. & James, R. Immunity proteins and their specificity for endonuclease colicins: telling right from wrong in protein-protein recognition. *Mol. Microbiol.* **28,** 227–233 (1998).
138. Kaur, S. & Kaur, S. Bacteriocins as Potential Anticancer Agents. *Front. Pharmacol.* **6,** 272 (2015).
139. Grinter, R., Milner, J. & Walker, D. Bacteriocins active against plant pathogenic bacteria. *Biochem. Soc. Trans.* **40,** 1498–502 (2012).
140. Schulz, S. *et al.* Broad and efficient control of major foodborne pathogenic strains of Escherichia coli by mixtures of plant-produced colicins. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E5454-60 (2015).
141. Rassam, P. *et al.* Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. *Nature* **523,** 333–336 (2015).
142. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–72 (2008).
143. Fan, S.-B. *et al.* Using pLink to Analyze Cross-Linked Peptides. *Curr. Protoc. Bioinformatics* **49,** 8.21.1-19 (2015).
144. Kosinski, J. *et al.* Xlink Analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **189,** 177–183 (2015).
145. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–12 (2004).
146. Cyril Dominguez, Rolf Boelens, and & Bonvin*, A. M. J. J. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. (2003).
147. Rollauer, S. E., Sooreshjani, M. A., Noinaj, N. & Buchanan, S. K. Outer membrane protein biogenesis in Gram-negative bacteria. *Philos. Trans. R. Soc. London B Biol. Sci.* **370,** (2015).
148. Tamm, L. K., Arora, A. & Kleinschmidt, J. H. Structure and assembly of beta-barrel membrane proteins. *J. Biol. Chem.* **276,** 32399–402 (2001).
149. Tommassen, J. Assembly of outer-membrane proteins in bacteria and mitochondria. *Microbiology* **156,** 2587–96 (2010).
150. Wülfing, C. & Plückthun, A. Protein folding in the periplasm of Escherichia coli. *Mol. Microbiol.* **12,** 685–92 (1994).
151. Fleming, K. G. A combined kinetic push and thermodynamic pull as driving forces for outer membrane protein sorting and folding in bacteria. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370,** (2015).
152. Wu, T. *et al.* Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli. *Cell* **121,** 235–45 (2005).
153. Malinverni, J. C. *et al.* YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. *Mol. Microbiol.* **61,** 151–64 (2006).
154. Knowles, T. J., Scott-Tucker, A., Overduin, M. & Henderson, I. R. Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nat. Rev. Microbiol.* **7,** 206–214 (2009).
155. O'Neil, P. K., Rollauer, S. E., Noinaj, N. & Buchanan, S. K. Fitting the Pieces of the

β-Barrel Assembly Machinery Complex. *Biochemistry* **54,** 6303–6311 (2015).
156. Han, L. *et al.* Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins. *Nat. Struct. Mol. Biol.* **23,** 192–197 (2016).
157. Misra, R., Stikeleather, R. & Gabriele, R. In vivo roles of BamA, BamB and BamD in the biogenesis of BamA, a core protein of the β-barrel assembly machine of Escherichia coli. *J. Mol. Biol.* **427,** 1061–74 (2015).
158. Knowles, T. J. *et al.* Structure and function of BamE within the outer membrane and the β-barrel assembly machine. *EMBO Rep.* **12,** 123–8 (2011).
159. Bakelar, J., Buchanan, S. K. & Noinaj, N. The structure of the β-barrel assembly machinery complex. *Science* **351,** 180–6 (2016).
160. Kim, K. H., Aulakh, S. & Paetzel, M. The bacterial outer membrane β-barrel assembly machinery. *Protein Sci.* **21,** 751–68 (2012).
161. Hagan, C. L., Wzorek, J. S. & Kahne, D. Inhibition of the β-barrel assembly machine by a peptide that binds BamD. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 2011–6 (2015).
162. Struyvé, M., Moons, M. & Tommassen, J. Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J. Mol. Biol.* **218,** 141–8 (1991).
163. de Cock, H., Struyvé, M., Kleerebezem, M., van der Krift, T. & Tommassen, J. Role of the carboxy-terminal phenylalanine in the biogenesis of outer membrane protein PhoE of Escherichia coli K-12. *J. Mol. Biol.* **269,** 473–8 (1997).
164. Robert, V. *et al.* Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol.* **4,** e377 (2006).
165. Lehr, U. *et al.* C-terminal amino acid residues of the trimeric autotransporter adhesin YadA of Yersinia enterocolitica are decisive for its recognition and assembly by BamA. *Mol. Microbiol.* **78,** 932–46 (2010).
166. Paramasivam, N. *et al.* Is the C-terminal insertional signal in Gram-negative bacterial outer membrane proteins species-specific or not? *BMC Genomics* **13,** 510 (2012).
167. Kozjak-Pavlovic, V., Ott, C., Götz, M. & Rudel, T. Neisserial Omp85 protein is selectively recognized and assembled into functional complexes in the outer membrane of human mitochondria. *J. Biol. Chem.* **286,** 27019–26 (2011).
168. Walther, D. M., Papic, D., Bos, M. P., Tommassen, J. & Rapaport, D. Signals in bacterial beta-barrel proteins are functional in eukaryotic cells for targeting to and assembly in mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 2531–6 (2009).
169. Jiang, J.-H., Davies, J. K., Lithgow, T., Strugnell, R. A. & Gabriel, K. Targeting of Neisserial PorB to the mitochondrial outer membrane: an insight on the evolution of β-barrel protein assembly machines. *Mol. Microbiol.* **82,** 976–87 (2011).
170. Jiang, J.-H., Tong, J., Tan, K. S. & Gabriel, K. From evolution to pathogenesis: the link between β-barrel assembly machineries in the outer membrane of mitochondria and gram-negative bacteria. *Int. J. Mol. Sci.* **13,** 8038–50 (2012).
171. Sharma, O. *et al.* Genome-wide screens: novel mechanisms in colicin import and cytotoxicity. *Mol. Microbiol.* **73,** 571–85 (2009).
172. Anderluh, G. *et al.* Concerted folding and binding of a flexible colicin domain to its periplasmic receptor TolA. *J. Biol. Chem.* **278,** 21860–8 (2003).
173. Hecht, O., Ridley, H., Lakey, J. H. & Moore, G. R. A common interaction for the entry of colicin N and filamentous phage into Escherichia coli. *J. Mol. Biol.* **388,** 880–93 (2009).
174. Gokce, I. *et al.* The TolA-recognition site of colicin N. ITC, SPR and stopped-flow fluorescence define a crucial 27-residue segment. *J. Mol. Biol.* **304,** 621–32 (2000).

175. Housden, N. G. *et al.* Directed epitope delivery across the Escherichia coli outer membrane through the porin OmpF. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21412–7 (2010).
176. Penfold, C. N. *et al.* Flexibility in the receptor-binding domain of the enzymatic colicin E9 is required for toxicity against Escherichia coli cells. *J. Bacteriol.* **186,** 4520–7 (2004).
177. Bruce, J. E. In vivo protein complex topologies: sights through a cross-linking lens. *Proteomics* **12,** 1565–75 (2012).
178. Weisbrod, C. R. *et al.* In vivo protein interaction network identified with a novel real-time cross-linked peptide identification strategy. *J. Proteome Res.* **12,** 1569–79 (2013).
179. Holding, A. N. XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods* **89,** 54–63 (2015).
180. Merkley, E. D., Cort, J. R. & Adkins, J. N. Cross-linking and mass spectrometry methodologies to facilitate structural biology: finding a path through the maze. *J. Struct. Funct. Genomics* **14,** 77–90 (2013).
181. Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173,** 530–40 (2011).
182. Farrell, I. S., Toroney, R., Hazen, J. L., Mehl, R. A. & Chin, J. W. Photo-cross-linking interacting proteins with a genetically encoded benzophenone. *Nat. Methods* **2,** 377–384 (2005).
183. Leitner, A. *et al.* Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **9,** 1634–49 (2010).
184. Leitner, A. *et al.* The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* **20,** 814–25 (2012).
185. Cline, G. W. & Hanna, S. B. Kinetics and mechanisms of the aminolysis of N-hydroxysuccinimide esters in aqueous buffers. *J. Org. Chem.* **53,** 3583–3586 (1988).
186. Okayasu, T., Ikeda, M., Akimoto, K. & Sorimachi, K. The amino acid composition of mammalian and bacterial cells. *Amino Acids* **13,** 379–391 (1997).
187. Kalkhof, S. & Sinz, A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal. Bioanal. Chem.* **392,** 305–312 (2008).
188. Chen, Z. A. *et al.* Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29,** 717–26 (2010).
189. Fritzsche, R., Ihling, C. H., Götze, M. & Sinz, A. Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. *Rapid Commun. Mass Spectrom.* **26,** 653–658 (2012).
190. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9,** 904–6 (2012).
191. Boetzel, R. *et al.* Structural dynamics of the receptor-binding domain of colicin E9. *Faraday Discuss.* **122,** 145-62-90 (2003).
192. Lotze, J., Reinhardt, U., Seitz, O. & Beck-Sickinger, A. G. Peptide-tags for site-specific protein labelling in vitro and in vivo. *Mol. BioSyst.* **12,** 1731–1745 (2016).